



(51) International Patent Classification <sup>6</sup> : <b>C12Q 1/68, C12N 15/00, A01N 37/18, A61K 38/00, 48/00, C07H 21/02</b>		<b>A1</b>	(11) International Publication Number: <b>WO 96/06188</b>
			(43) International Publication Date: 29 February 1996 (29.02.96)
(21) International Application Number: <b>PCT/US95/10523</b>		(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).	
(22) International Filing Date: 17 August 1995 (17.08.95)			
(30) Priority Data: 292,902 18 August 1994 (18.08.94) US 515,190 15 August 1995 (15.08.95) US			
(71) Applicant: CYTOGEN CORPORATION [US/US]; 600 Col- lege Road East, Princeton, NJ 08540 (US).		Published With international search report.	
(72) Inventors: RODWELL, John, D.; 1340 Eagle Road, New Hope, PA 18938 (US). FOWLKES, Dana, M.; 2013 Damascus Church Drive, Chapel Hill, NC 27516 (US).			
(74) Agents: MISROCK, S., Leslie et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).			
(54) Title: PEPTIDE LIBRAIRIES AS A SOURCE OF SYNGENES			
(57) Abstract <p>The present invention relates generally to synthetic gene sequences ("syngenes") that are identified by screening random peptide libraries for peptides that bind a ligand of choice. The synthetic gene sequences, together with, optionally, other DNA sequences that target the synthetic gene sequences or their encoded proteins to particular locations <i>in vivo</i> or intracellularly, or that contain processing signals, or that code for other peptides or amino acid sequences, are cloned into suitable expression vectors. The syngenes are used, for example, in gene therapy to supply, via expression of their encoded proteins, a therapeutic product. In another aspect, the invention relates to protein or peptide products of syngenes and their therapeutic and diagnostic uses.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## PEPTIDE LIBRARIES AS A SOURCE OF SYNGENES

This application is a continuation-in-part of co-pending U.S. Patent Application Serial No. 08/292,902 filed August 18, 1994, the entire disclosure of which is incorporated herein by reference.

### 1. FIELD OF THE INVENTION

The present invention relates generally to synthetic gene sequences ("syngenes"), particularly for use in gene therapy. Syngenes are nucleic acids that comprise synthetic gene sequences identified by screening synthetic random peptide libraries, for peptides that bind a ligand of choice. The synthetic gene sequences, together with, optionally, other DNA sequences that target the synthetic gene sequences or their encoded proteins to particular locations in vivo or intracellularly, or that contain processing signals, or that code for other peptides or amino acid sequences, are cloned into suitable expression vectors. The syngenes are used, for example, in gene therapy to supply, via expression of their encoded proteins, a therapeutic product. In another aspect, the invention relates to protein or peptide products of syngenes and their therapeutic and diagnostic uses.

### 2. BACKGROUND OF THE INVENTION

#### 2.1. GENE THERAPY

Gene therapy today is often directed to the replacement of a product of a defective gene. For example, a human adenosine deaminase gene has been engineered into skin fibroblasts of mice through the use of a retroviral vector (Palmer et al., 1988, Proc. Natl. Acad. Sci. USA 88:1330-1334). The human  $\beta$ -globin gene has been transferred into mouse bone marrow cells and mouse erythroleukemia cells as a

model for the treatment of  $\beta$ -thalassemia and sickle cell anemia (Novak et al., 1990, Proc. Natl. Acad. Sci. USA 87:3386-3390; Bender et al., 1988, Mol. Cell Biol. 8:1725-1735; Chao et al., 1983, Cell 32:483-493).

Other examples of the replacement of a defective gene product include: the delivering of the human cystic fibrosis transmembrane conductance regulator via a retrovirus into airway epithelia of mice (Hyde et al., 1993, Nature 362:250-255); in a rabbit model, the transfer of human low density lipoprotein receptor cDNA into hepatocytes (Wilson et al., 1992, J. Biol. Chem. 267:963-967); the transfer of human Factor IX cDNA into skin fibroblasts, followed by engraftment of the fibroblasts into mice (Louis and Verma, 1988, Proc. Natl. Acad. Sci. USA 85:3150-3154).

In addition to replacing a defective gene product, gene therapy can also be used to provide a cell or an organism with a new function. Examples of such a use include: transferring the human growth hormone gene via a retrovirus into human keratinocytes and then grafting the keratinocytes into nude mice. (Morgan et al., 1987, Science 237:1476-1479); expressing therapeutics for cardiovascular disease in endothelial cells with the aim of delivering the therapeutics through the circulation (Nabel et al., 1989, Science 244:1342-1344); the use of antisense c-myc oligonucleotides to prevent the proliferation of vascular smooth muscle cells (Simons et al., 1992, Nature 359:67-70).

Another aspect of gene therapy involves the inhibition or enhancement of the activity of a preselected gene. The preselected gene is one that is involved in some aspect of a diseased state. The ability to modulate the activity of such a gene would be useful in the treatment of the diseased state. In

this aspect, gene therapy is limited by the availability of specific modulators of specific target genes. Currently, there is no easy method of increasing the specificity of known modulators or  
5 discovering new modulators. It would be of great value to have reagents and methods for identifying such modulators. This would permit virtually any gene to be the subject of this aspect of gene therapy.

Currently, all gene therapy programs use  
10 either 1) cloned genes that are endogenous in some organism or 2) genes encoding mRNAs that are antisense to known genes. Current methods of gene therapy usually utilize human genes, either in their sense or anti-sense orientations. It would be of great value  
15 to have a source of genes other than the human genome since the use of the human genome presents certain problems. Using the human genome entails the laborious procedures of the identification and isolation of a gene having a desired function. A  
20 further problem is the possibility that no human gene may possess the desired function. Even if a human gene is found that has the desired function, that gene may not have the desired specificity; it may be too large to be easily used in current gene therapy  
25 protocols; it may be encoded by multiple exons; and fragments of the gene may not be suitable for use in gene therapy because the gene's binding regions may be too complex for fragments of the endogenous protein to mimic its function. In order to avoid these problems,  
30 it would be highly desirable to have a method of producing genes for use in gene therapy that does not rely on the human genome as a source of those genes.

## 2.2. PEPTIDE LIBRARIES

35 The use of peptide libraries is well known in the art. Such peptide libraries have generally been constructed by one of two approaches. According

to one approach, peptides have been chemically synthesized *in vitro* in several formats. For example, Fodor, S., et al., 1991, Science 251: 767-773, describes use of complex instrumentation, photochemistry and computerized inventory control to synthesize a known array of short peptides on an individual microscopic slide. Houghten, R., et al., 1991, Nature 354: 84-86, describes mixtures of free hexapeptides in which the first and second residues in each peptide were individually and specifically defined. Lam, K., et al., 1991, Nature 354: 82-84, describes a "one bead, one peptide" approach in which a solid phase split synthesis scheme produced a library of peptides in which each bead in the collection had immobilized thereon a single, random sequence of amino acid residues. For the most part, the chemical synthetic systems have been directed to generation of arrays of short length peptides, generally fewer than about 10 amino acids or so, more particularly about 6-8 amino acids. Direct amino acid sequencing, alone or in combination with complex record keeping of the peptide synthesis schemes, is required to use these libraries.

According to a second approach using recombinant DNA techniques, peptides have been expressed in biological systems as either soluble fusion proteins or viral capsid fusion proteins.

A number of peptide libraries according to the second approach have used the M13 phage. M13 is a filamentous bacteriophage that has been a workhorse in molecular biology laboratories for the past 20 years. M13 viral particles consist of six different capsid proteins and one copy of the viral genome, as a single-stranded circular DNA molecule. Once the M13 DNA has been introduced into a host cell such as *E. coli*, it is converted into double-stranded, circular DNA. The viral DNA carries a second origin of

replication that is used to generate the single-stranded DNA found in the viral particles. During viral morphogenesis, there is an ordered assembly of the single-stranded DNA and the viral proteins, and the viral particles are extruded from cells in a process much like secretion. The M13 virus is neither lysogenic nor lytic like other bacteriophage (e.g.,  $\lambda$ ); cells, once infected, chronically release virus. This feature leads to high titers of virus in infected cultures, i.e.,  $10^{12}$  pfu/ml.

The genome of the M13 phage is ~8000 nucleotides in length and has been completely sequenced. The viral capsid protein, protein III (pIII) is responsible for infection of bacteria. In *E. coli*, the pillin protein encoded by the F factor interacts with pIII protein and is responsible for phage uptake. Hence, all *E. coli* hosts for M13 virus are considered male because they carry the F factor. Several investigators have determined from mutational analysis that the 406 amino acid long pIII capsid protein has two domains. The C-terminus anchors the protein to the viral coat, while portions of the N-terminus of pIII are essential for interaction with the *E. coli* pillin protein (Crissman, J.W. and Smith, G.P., 1984, *Virology* 132: 445-455). Although the N-terminus of the pIII protein has been shown to be necessary for viral infection, the extreme N-terminus of the mature protein does tolerate alterations. In 1985, George Smith published experiments reporting the use of the pIII protein of bacteriophage M13 as an experimental system for expressing a heterologous protein on the viral coat surface (Smith, G.P., 1985, *Science* 228: 1315-1317). It was later recognized, independently by two groups, that the M13 phage pIII gene display system could be a useful one for mapping antibody epitopes. De la Cruz, V., et al., (1988, *J. Biol. Chem.* 263: 4318-4322) cloned and expressed

segments of the cDNA encoding the *Plasmodium falciparum* surface coat protein into the pIII gene, and recombinant phage were tested for immunoreactivity with a polyclonal antibody. Parmley, S.F. and Smith, G.P., (1988, Gene 73: 305-318) cloned and expressed segments of the *E. coli*  $\beta$ -galactosidase gene in the pIII gene and identified recombinants carrying the epitope of an anti- $\beta$ -galactosidase monoclonal antibody. The latter authors also described a process termed "biopanning", in which mixtures of recombinant phage were incubated with biotinylated monoclonal antibodies, and phage-antibody complexes could be specifically recovered with streptavidin-coated plastic plates.

In 1989, Parmley and Smith, 1989, Adv. Exp. Med. Biol. 251:215-218 suggested that short, synthetic DNA segments cloned into the pIII gene might represent a library of epitopes. These authors reasoned that since linear epitopes were often ~6 amino acids in length, it should be possible to use a random recombinant DNA library to express all possible hexapeptides to isolate epitopes that bind to antibodies.

Scott and Smith, 1990, Science 249:386-390 describe construction and expression of an "epitope library" of hexapeptides on the surface of M13. The library was made by inserting a 33 base pair Bgl I digested oligonucleotide sequence into an Sfi I digested phage fd-tet, i.e., fUSE5 RF. The 33 base pair fragment contains a random or "degenerate" coding sequence (NNK), (SEQ ID NO: 1) where N represents G, A, T or C and K represents G or T. The authors stated that the library consisted of  $2 \times 10^8$  recombinants expressing  $4 \times 10^7$  different hexapeptides; theoretically, this library expressed 69% of the  $6.4 \times 10^7$  possible peptides ( $20^6$ ). Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87: 6378-6382 also



d scribed a somewhat similar library of hexapeptides expressed as pIII gene fusions of M13 fd phage. PCT publication WO 91/19818 dated December 26, 1991 by Dower and Cwirla describes a similar library of  
5 pentameric to octameric random amino acid sequences.

Devlin et al., 1990, Science, 249:404-406, describes a peptide library of about 15 residues generated using an (NNS) coding scheme for oligonucleotide synthesis in which S is G or C.

10 Christian et al., 1992, J. Mol. Biol. 227:711-718 described a phage display library, expressing decapeptides. The starting DNA was generated by means of an oligonucleotide comprising the degenerate codons [NN(G/T)]<sub>10</sub> (SEQ ID NO: 2) with a  
15 self-complementary 3' terminus. This sequence, in forming a hairpin, creates a self-priming replication site which could be used by T4 DNA polymerase to generate the complementary strand. The double-stranded DNA was cleaved at the SfiI sites at the 5'  
20 terminus and hairpin for cloning into the fUSE5 vector described by Scott and Smith, *supra*.

Other investigators have used other viral capsid proteins for expression of non-viral DNA on the surface of phage particles. The protein pVIII is a  
25 major M13 viral capsid protein and interacts with the single stranded DNA of M13 viral particles at its C-terminus. It is 50 amino acids long and exists in approximately 2,700 copies per particle. The N-terminus of the protein is exposed and will tolerate  
30 insertions, although large inserts have been reported to disrupt the assembly of pVIII fusion proteins into viral particles (Cesareni, 1992, FEBS Lett. 307:66-70). To minimize the negative effect of pVIII fusion proteins, a phagemid system has been utilized.  
35 Bacterial cells carrying the phagemid are infected with helper phage and secrete viral particles that have a mixture of both wild-type and pVIII fusion

capsid molecules. pVIII has also served as a site for expressing peptides on the surface of M13 viral particles. Four and six amino acid sequences corresponding to different segments of the *Plasmodium falciparum* major surface antigen have been cloned and expressed in the comparable gene of the filamentous bacteriophage fd (Greenwood et al., 1991, J. Mol. Biol. 220:821-827).

Lenstra, 1992, J. Immunol. Meth. 152:149-157 describes construction of a library by a laborious process encompassing annealing oligonucleotides of about 17 or 23 degenerate bases with an 8 nucleotide long palindromic sequence at their 3' ends. This resulted in the expression of random hexa- or octapeptides as fusion proteins with the  $\beta$ -galactosidase protein in a bacterial expression vector. The DNA was then converted into a double-stranded form with Klenow DNA polymerase, blunt-end ligated into a vector, and then released as Hind III fragments. These fragments were then cloned into an expression vector at the C-terminus of a truncated  $\beta$ -galactosidase to generate 10' recombinants. Colonies were then lysed, blotted on nitrocellulose filters (10<sup>4</sup>/filter) and screened for immunoreactivity with several different monoclonal antibodies. A number of clones were isolated by repeated rounds of screening and were sequenced.

Screening of peptide libraries has generally been confined to the use of a restricted number of ligands. Most commonly, the ligand has been an antibody (Parmley and Smith, 1989, Adv. Exp. Med. Biol. 251:215-218; Scott and Smith, 1990, Science 249:386-390). Streptavidin (Fowlkes et al., 1992; BioTechniques, 13:422-427) and concanavalin A (Oldenburg et al., 1992, Proc. Natl. Acad. Sci. USA 89:5393-5397) have also been used as ligands. SH2 and SH3 domains have been used (Yu et al., 1994, Cell 76:933-945).

Oligonucleotides have been used to screen a conventional cDNA library (Staudt et al., 1988, Science 241: 577-580). This resulted in the isolation of naturally occurring DNA binding proteins. Also, random DNA libraries have been screened for ligand binding properties (Bock et al., 1992, Nature 355:564-566; Tuerk et al., 1992, Proc. Natl. Acad. Sci. USA 89:6988-6992; Ellington et al., 1992, Nature 355:850-852). U. S. Patent No. 5,096,815, U.S. Patent No. 5,223,409, and U.S. Patent No. 5,198,346, all to Ladner et al., disclose the use of oligonucleotide ligands to screen a phage display library in which known, naturally occurring DNA binding proteins have been cloned. Following mutagenesis of certain defined positions in the sequences of the naturally occurring DNA binding proteins, stronger binding versions of those proteins were isolated. Another example of the screening of a non-random phage display library with an oligonucleotide is given in Rebar and Pabo, 1993, Science 263:671-673. This work resulted in the isolation of variants of known zinc finger DNA binding proteins. Rebar and Pabo do not disclose the use of random peptide libraries that are totally synthetic. As yet, oligonucleotides have not been used to screen a totally synthetic, random, phage display peptide library.

A common goal of screening peptide libraries has generally been to find a peptide with a desired biological effect and (1) use that peptide directly; (2) use the peptide as a basis for designing peptidomimetics for therapeutic use; and (3) use the peptide for research to map protein/target interactions, such as epitope mapping. Under current practice, once an appropriate peptide is identified from a library, that peptide, or a peptidomimetic of it, may be used therapeutically. However, use of such peptides or peptidomimetics may lead to problems of

peptide instability, delivery problems, or difficulty in making the peptidomimetic. The use of the syngenes of the present invention can obviate such difficulties. The use of syngenes is especially  
5 advantageous in cases where use of a long peptide (>20 amino acids) is necessary. In such cases, making an appropriate peptidomimetic is difficult or impossible.

U.S. Patent 5,198,346 to Ladner et al. ("Ladner"), at column 20, suggests the *in vivo*  
10 therapeutic use of genes encoding certain DNA binding proteins selected from peptide display libraries. Ladner does not, however, disclose the use of totally synthetic random peptide libraries. In contrast to  
15 the syngenes of the present invention, Ladner stresses the use of peptide display libraries that are derived from naturally occurring sequences (in this instance, sequences encoding known binding proteins) by *in vitro* mutagenesis of these sequences. Prior to the present  
20 invention, it had not been recognized that totally synthetic random peptide libraries are a tremendous source of synthetic genes.

Citation or identification of any reference herein shall not be construed as an admission that such reference is available as prior art to the  
25 present invention.

### 3. SUMMARY OF THE INVENTION

The present invention relates generally to synthetic gene sequences ("syngenes"), particularly  
30 for use in gene therapy. Syngenes are nucleic acids that comprise synthetic gene sequences identified by screening random peptide libraries for peptides that bind a ligand of choice. The synthetic gene sequences are cloned into suitable expression vectors together  
35 with, optionally, other DNA sequences that target the synthetic gene sequences or their encoded proteins to particular locations *in vivo* or intracellularly, or

that contain processing signals, or that code for other peptides or amino acid sequences. The syngenes are used, for example, in gene therapy to supply a therapeutic product via expression of their encoded proteins. In another aspect, the invention relates to protein or peptide products of syngenes and their therapeutic and diagnostic uses.

Pharmaceutical compositions comprising syngenes or their encoded peptides are also provided.

In one embodiment, the invention provides methods of identifying syngenes that bind to a particular ligand of choice. In specific aspects, the ligand of choice is a transcriptional regulatory site, a transcription factor that binds to a transcriptional regulatory site, or a binding partner/inhibitor of the transcription factor. In one aspect, such a transcriptional regulatory site is an NF- $\kappa$ B, AP-1, or ATF binding site. In other aspects, the transcription factor is NF- $\kappa$ B, AP-1, or ATF. Accordingly, in a specific embodiment, the present invention provides methods for identifying syngenes that inhibit or enhance the transcriptional activity of a wide variety of naturally occurring genes, preferably with a specificity not found in natural systems. Syngenes are also useful for modulating signal transduction pathways, metabolic pathways, RNA translation, and intracellular trafficking. In the cell membrane, syngenes may be used to modulate the activity of membrane receptors, ion channels, or exocytotic and endocytotic pathways. In tissue, syngenes, via expression of their encoded proteins, may be used to regulate cell/cell signalling and transcytosis. Cell/cell junctions and the extracellular matrix are appropriate targets for syngenes. Syngenes may be used to regulate cell adhesion or cell/cell recognition. In the general circulation, syngenes may be used to regulate the activity of receptor ligands.

#### 4. FIGURE LEGENDS

The present invention may be understood more fully by reference to the following detailed description of the invention, examples of specific  
5       embodiments of the invention and the appended figures in which:

Figure 1 (A-F) schematically illustrates construction of TSAR libraries. Figure 1A  
10       schematically depicts the synthesis and assembly of synthetic oligonucleotides for the linear libraries and bimolecular libraries illustrated in Figure 1B and 1C.  $N = A, C, G \text{ or } T$ ;  $B = G, T \text{ or } C$  and  $V = G, A, \text{ or } C$ ; and  $n$  and  $m$  are integers, such that  $10 \leq n \leq 100$  and  $10 \leq m \leq 100$ . Figure 1D-F schematically depicts  
15       representative libraries which are designed to be semirigid libraries. The synthesis and assembly of the oligonucleotides for the semirigid libraries are as in Figure 1A with modifications to include specified invariant positions. See Section 5.1.3.1.

20       Figure 2 schematically depicts an exemplary mRNA expressed by a syngene.

Figure 3 is a schematic depiction of a shuttle vector which can be used in one embodiment of the invention. (1) and (2) allow replication in *E.*  
25       coli and mammalian cells, respectively. (3) allows selection in *E. coli*. (4) and (5) allow transcription and mRNA processing, respectively, in mammalian cells. (6) is a syngene that preferably encodes amino-terminal spacer sequences, binding domain,  
30       targeting/localization signal, and, optionally, a second functional domain.

Figure 4 presents the results of ELISAs on phage clones isolated from the R26 library using the H2kB oligonucleotide as a target. See Section 6.1.4  
35       for details. Numbers along the abscissa refer to phage clones as follows: 1 = clone 1; 2 = clone 2; 3 = clone 6; 4 = clone 5; 5 = m663; 6 = a random clone.

H2κB, IL6κB, IL8κB, and NFIL6 refer to oligonucleotide ligands as described in Section 6.1.1. Phage clones were tested for binding to plates coated with the oligonucleotide ligands. Phage clones were also  
 5 tested for binding to plates coated with BSA. The height of the bars above the numbers along the abscissa represents the ratio of a phage clone's binding to an oligonucleotide ligand-coated plate compared to the phage clone's binding to the BSA-coated plate.  
 10

Figure 5 schematically illustrates construction of the TSAR-9 library. N = A, C, G or T; B = G, T or C and V = G, A or C.  
 GGCTCGAGN(NNB)<sub>10</sub>CCAGGT is SEQ ID NO: 3;  
 15 GGTCTAGA(VNN)<sub>10</sub>ACCTGG is SEQ ID NO: 4;  
 TCGAGN(NNB)<sub>10</sub>CCAGGT is SEQ ID NO: 5;  
 CTAGA(VNN)<sub>10</sub>ACCTGG is SEQ ID NO: 6;  
 SHSS(R/S)<sub>X10</sub>PGX<sub>10</sub>SRPSRT is SEQ ID NO: 7.

Figure 6 schematically illustrates the m663 expression vector.  
 20

Figure 7 schematically illustrates construction of the TSAR-12 library. N = A, C, G or T; B = G, T or C and V = G, A or C. Insertion into a representative, appropriate vector and expression in  
 25 an appropriate host is illustrated.  
 ttttgtcgacN(NNB)<sub>10</sub>Ngcggtg is SEQ ID NO: 8;  
 ttttactagt(VNN)<sub>10</sub>VNcaccgc is SEQ ID NO: 9;  
 tcgacN(NNB)<sub>10</sub>Ngcggtg is SEQ ID NO: 10;  
 ctagt(VNN)<sub>10</sub>VNcaccgc is SEQ ID NO: 11.

Figure 8 schematically illustrates construction of the TSAR-13 library.  
 TGACGACTCGAGTTGTGGT(NNK)<sub>10</sub>GGTTGTGGA(NNK)<sub>10</sub>GGGTGCGGC is  
 SEQ ID NO: 12; GATCCTTCTAGAACCTGGAGGCCACAGCC(MNN)<sub>10</sub>  
 GCCGCACCC is SEQ ID NO: 13; TCGAGTTGTGGT(NNK)<sub>10</sub>GGTTGT  
 35 GGA(NNK)<sub>10</sub>GGGTGCGGC is SEQ ID NO: 14; CTAGAACCTGGAGGC  
 CCACAGCC(MNN)<sub>10</sub>GCCGCACCC is SEQ ID NO: 15;  
 SSCGX<sub>10</sub>GCGX<sub>10</sub>GCGX<sub>10</sub>GCGPPGSR is SEQ ID NO: 16.

Figure 9 schematically depicts the construction of the R26 library. The R26 expression library was constructed essentially as described for the TSAR-9 library that is described in Section 6.9.1 and its subsections, except for the modifications depicted in Figure 9. ctgtgcctcgagB(NNB)<sub>1,2</sub>Nccgcgg is SEQ ID NO: 17; ctgtgctctaga(VNN)<sub>1,2</sub>VNccgcgg is SEQ ID NO: 18; tcgagB(NNB)<sub>1,2</sub>Nccgcgg is SEQ ID NO: 19; ctaga(VNN)<sub>1,2</sub>VNccgcgg is SEQ ID NO: 20;

10 SHSS(S/R)X<sub>1,2</sub>πAδX<sub>1,2</sub>SRPSRT is SEQ ID NO: 21.

Figure 10 (A-D) represents circular restriction maps of phagemid vectors, derived from phagemid pBluescript II SK<sup>+</sup>, in which a truncated portion encoding amino acid residues 198-406 of the pIII gene of M13 is linked to a leader sequence of the E. coli Pel B gene and is expressed under control of a lac promoter. G and S represent the amino acids glycine and serine, respectively; c-myc represents the human c-myc oncogene epitope recognized by the 9E10 monoclonal antibody described in Evan et al., 1985, Mol. Cell. Biol. 5:3610-3616. Figure 10A illustrates the restriction map of phagemid pDAF1; Figure 10B illustrates the restriction map of phagemid pDAF2; Figure 10C illustrates the restriction map of phagemid pDAF3; Figure 10D schematically illustrates the construction of phagemids pDAF1, pDAF2 and pDAF3.

25

Figure 11 schematically depicts the construction of the R8C library. The R8C expression library was constructed essentially as described for the TSAR-9 library that is described in Section 6.9.1 and its subsections, except for the modifications depicted in Figure 11.

30

TGACGTCTCGAGTTGT(NNK)<sub>1</sub>TGTGGATCTAGAAGGATC is SEQ ID NO: 22; GATCCTTCTAGATCC is SEQ ID NO: 23;

35 TCGAGTTGT(NNK)<sub>1</sub>TGTGGAT is SEQ ID NO: 24; TCAGATCCACA(MNN)<sub>1</sub>ACAAC is SEQ ID NO: 25; SSCX<sub>1</sub>CGSRPSRT is SEQ ID NO: 26.



Figure 12 schematically depicts the origin of one class of double-insert recombinants in the R8C library. TCGAGTTGT(NNK)<sub>3</sub>TGTGGATCTAGATCCACA(MNN)<sub>3</sub>ACAAC is SEQ ID NO: 27; TCGAGTTGT(NNK)<sub>3</sub>TGTGGA

- 5 TCTAGATCCACA(MNN)<sub>3</sub>ACAAC is SEQ ID NO: 28;  
SSCX<sub>3</sub>CGSRSTX<sub>3</sub>TTR is SEQ ID NO: 29.

Figure 13 schematically illustrates the construction of the DC43 TSAR library. The DC43 expression library was constructed essentially as described for the TSAR-9 library that is described in Section 6.9.1 and its subsections, except for the modifications depicted in Figure 13.

- 10 GTGTGTCTCGAGN(NNB)<sub>20</sub>GGTTGTGGT is SEQ ID NO: 30;  
GTTGTGTCTAGA(VNN)<sub>20</sub>ACCACAACC is SEQ ID NO: 31;  
15 TCGAGN(NNB)<sub>20</sub>GGTTGTGGT is SEQ ID NO: 32;  
CTAGA(VNN)<sub>20</sub>ACCACAACC is SEQ ID NO: 33;  
HSS(S/R)<sub>X20</sub>GCGX<sub>20</sub>SRIEGRARPSR is SEQ ID NO: 34.

Figure 14 shows the selectivity of binding of two H2κB binding phage (H2κB-1 and H2κB-2), the NFIL6 binding phage NFIL6-1, and the parental phage m663 for three different target sites: the H2κB oligonucleotide, the NFIL6 oligonucleotide, and the IL6κB oligonucleotide. The numbers along the abscissa refer to the various phage as follows: 1 = H2κB-1; 2 =

20 H2κB-2; 3 = m663; 4 = NFIL6-1.

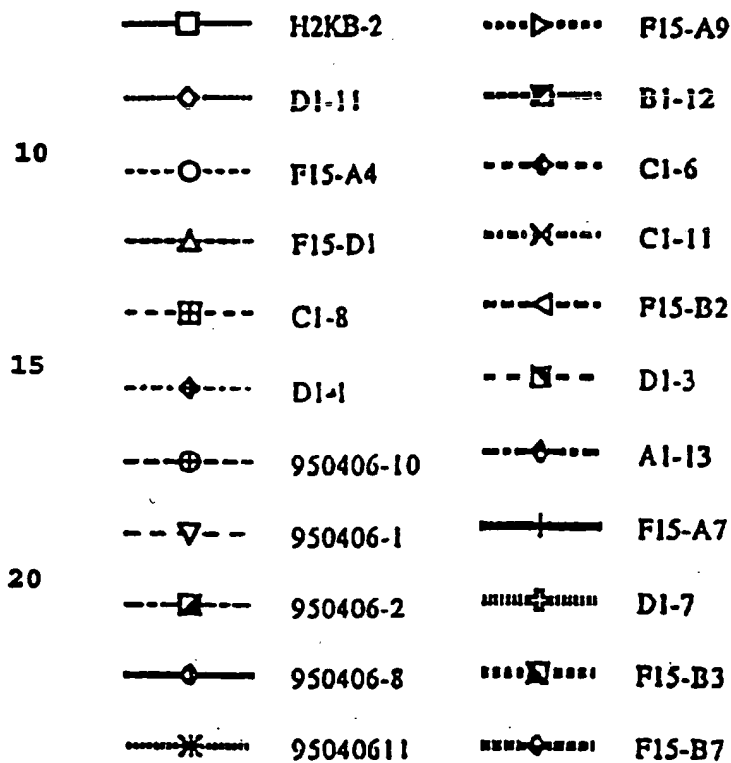
Figure 15 illustrates the molecular evolution scheme for phage H2κB-2 described in Section 6.1.5 that produced the ME#1 library. \*p indicates the approximate frequency at which the original amino acid residue is expected to occur in the phage of the ME#1 library. SSRTGNEQPPGSFGRAGCFHPGCKYMKLN is SEQ ID NO: 35; TCCTCGAGaacbgbgbaabgabcabccbccbggbtcbbtbgg bcgbgcbnCTGGTTnbtbcbabccbggbtgbaabtabatbaabctbaab is SEQ ID NO: 36; GTGTGTGTCTCGAGaacbgbgbaabgabcabccbccbgg btbcbtbggbcggbgcbnCTGGTTn is SEQ ID NO: 37; GTGTGTGTCT CGAvttvagvttvatvtavttvcavccvggvtgvaavnAACCAGn is SEQ ID NO: 38.

30

35

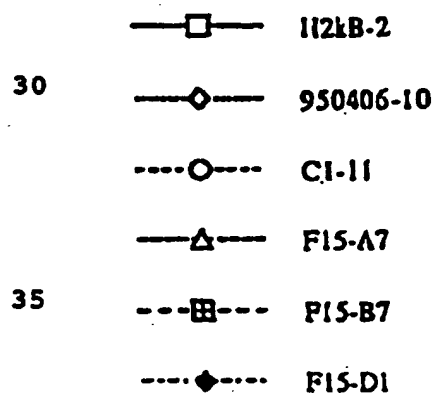
Figures 16A and 16B present the results of phage ELISAs that show the relative binding avidity for the H2kB oligonucleotide of clones isolated from the MB#1 library as compared to the H2kB-2 clone.

5 For 16A:



25

For 16B:



35

Figures 17A and 17B illustrate the molecular evolution schemes described in Section 6.1.5 that produced the ME#2a and ME#2b libraries, respectively.

5 TGTGTGTGTTCTCGAGN(NNB)<sub>10</sub>CGTTCCGGTCGTGGTTGCTTCCACCCGGGT  
 TGCAAAUACATCAAACACAAA(NNB)<sub>10</sub>TCTAGAACACACACA is SEQ ID NO: 39; TGTGTGTGTTCTAGA(VNN)<sub>10</sub>TTTGTGTTTGATGTATTTGCAACCCGGGTGGAAGCAACCACGACCGGAACG(VNN)<sub>10</sub>NCTCGAGGAACACACACA is SEQ ID NO: 40; SS(S/R)X<sub>12</sub>RSGRGCFHPGCKYIKHKX<sub>10</sub>SR is SEQ ID NO: 41; TGTGTGTGTTCTAGA(VNN)<sub>10</sub>TTTGTGTTTGATGTATTTGCAACCCGGGTGGAAGC is SEQ ID NO: 42; TGTGTGTGTTCTCGAGN(NNB)<sub>10</sub>CGTTCCGGTCGTGGTTGCTTCCACCCGGG is SEQ ID NO: 43; TGTGTCAACCTCTCGAGN(NNB)<sub>12</sub>GGTTGCTTCCACCCGGGTGCAAAATACATCAAACACAAG(NNB)<sub>12</sub>TCTAGAACACACACA is SEQ ID NO: 44; TGTGTGTGTTCTAGA(VNN)<sub>12</sub>CTTGTGTTTGATGTATTTGCAACCCGGTGGAAAGCAAC(VNN)<sub>12</sub>NCTCGAGAGGAACACACACA is SEQ ID NO: 45; SS(S/R)X<sub>12</sub>GCFHPGCKYIKHKX<sub>12</sub>SR is SEQ ID NO: 46; TGTGTGTGTTCTAGA(VNN)<sub>12</sub>CTTGTGTTTGATGTATTTGCAACCC is SEQ ID NO: 47; TGTGTGTGTTCTCGAGN(NNB)<sub>12</sub>GGTTGCTTCCACCCGGGTGCAAA is SEQ ID NO: 48.

20 Figure 18 presents the results of phage ELISAs that show the relative binding avidity for the H2kB oligonucleotide of clones isolated from the ME#2a and ME#2b libraries as compared to the H2kB-2 clone.

25	—□—	J5-1A7	--▽--	J5-1B6
	—◇—	J5-2A7	---■---	J5-1B1
	----○----	J5-1A10	—◇—	J5-2B3
30	—△—	J5-2A4	—*—	J5-2B2
	---■---	J5-2A8	....▷....	J5-2B4
	----◇----	J5-2B10	---■---	950406-10
35	---⊕---	J5-1B2	---◇---	H2kB-2

## 5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to nucleic acids ("syngenes") that comprise a random synthetic nucleic acid sequence that encodes a protein, peptide, or polypeptide (used interchangeably hereinafter and most commonly collectively referred to as "peptide" unless indicated otherwise explicitly or by context) that binds to a ligand of choice, and their uses in gene therapy for myriad diseases and disorders of interest. The invention provides methods for identification of syngenes that encode peptides that specifically bind to a ligand of choice. Also provided are compositions comprising syngenes as well as uses of the syngenes, e.g., in diagnosis and therapy of various disorders.

Syngenes encode a syngene product which is a peptide having at least one functional domain. The functional domain is a binding domain with affinity for a ligand of choice.

The syngenes of the present invention are synthetic genes, that is, genes that are not known to be present in a naturally occurring genome. In a specific embodiment, a syngene is a nucleic acid encoding a peptide comprising a binding domain in which the binding domain sequence is identified from a peptide library comprising at least 5 unpredictable contiguous amino acids in the variable portion of the library. In another embodiment, syngenes are made up of, at least in part, combinations of genes encoding functional domains, such combinations not occurring in nature. Syngenes may be composed of totally synthetic gene sequences, combinations of natural and totally synthetic gene sequences, or combinations of natural DNA sequences juxtaposed so as not to form a gene present in nature.

The present invention vastly expands the number of genes that are available for use in gene therapy. The present invention provides methods for finding synthetic nucleic acids encoding peptides with  
5 diagnostic or therapeutic value. The present invention further provides methods for delivery of such peptides *in vivo* by expression *in vivo* from the administered syngene, that potentially avoid the common problems of instability, clearance, etc. faced  
10 when dosing with proteins directly. More importantly, however, there are many instances where a protein that possesses a desired biological effect is unknown. For example, a known protein may not possess a desired level of binding specificity for a particular ligand.  
15 Syngenes encoding peptides with such novel specificities can be advantageously identified by the methods of the present invention.

The present invention provides methods for identifying synthetic peptides, and the nucleic acids  
20 encoding them, to fulfill roles that naturally occurring proteins have not necessarily been evolved for. One need only identify a target whose activity it would be desirable to modulate. Given such a target, a synthetic peptide in a random peptide  
25 library that binds the target and can thus modulate the target's activity can be readily identified by the present invention. The present invention also provides for the identification of a nucleic acid (syngene) that encodes the peptide. Such a gene,  
30 referred to as a syngene, can then be used, e.g., in gene therapy. Alternatively, the syngene may be introduced into an appropriate host cell and thereby used for the recombinant production of its encoded protein.

35 In a specific embodiment, the present invention provides methods for identifying syngenes that inhibit or enhance the transcriptional activity

of a wide variety of naturally occurring genes, preferably with a specificity not found in natural systems.

In addition to affecting transcription, among the other nuclear processes that syngenes may be useful in modulating are: post-transcriptional processing of RNA, DNA repair, and DNA replication.

Syngenes are also useful for modulating processes that occur outside of the nucleus. For example, in the cytoplasm, the encoded protein of a syngene, produced by intracellular expression of the syngene, may be used to affect the activity of inhibitors of transcription factors such as IF- $\kappa$ B. Syngenes may also be used to modulate signal transduction pathways, metabolic pathways, RNA translation, and intracellular trafficking. In the cell membrane, syngenes may be used to modulate the activity of membrane receptors, ion channels, or exocytotic and endocytotic pathways.

In tissue, syngenes, via expression of their encoded proteins, may be used to regulate cell/cell signalling and transcytosis. Cell/cell junctions and the extracellular matrix are appropriate targets for syngene expressed peptides. Syngenes may be used to regulate cell adhesion or cell/cell recognition. In the general circulation, syngenes may be used to regulate the activity of receptor ligands.

In another embodiment, the invention provides the peptides comprising binding domains that are encoded by syngenes, as well as their therapeutic and diagnostic uses, and compositions comprising such peptides.

#### 5.1. RANDOM PEPTIDE LIBRARIES FOR USE IN IDENTIFYING SYNTHETIC GENE SEQUENCES ENCODING A BINDING DOMAIN

Binding domains encoded by the syngenes of the invention can be identified from a random peptide expression library or a chemically synthesized random peptide library. A nucleic acid which expresses a peptide which binds to a ligand of choice can be identified and recovered from a random peptide expression library, and then sequenced to determine its nucleotide sequence and hence its deduced amino acid sequence that mediates binding. Alternatively, the amino acid sequence of an appropriate binding domain can be determined by direct determination of the amino acid sequence of a peptide selected from a random peptide library containing chemically synthesized peptides, whereby an appropriate syngene encoding the peptide can be designed and made. In a less preferred aspect, direct amino acid sequencing of a binding peptide selected from a random peptide expression library can also be performed, and an encoding nucleic acid designed. Where it is desired to decrease the size of the syngene-encoded binding domain, methods can be used to identify portions of the determined synthetic amino acid or nucleotide sequences which respectively mediate binding, or encode the sequences which mediate binding, as described in Section 5.3 below.

The term "random" peptide libraries is meant to include within its scope libraries of both partially and totally random peptides. Thus, peptide sequences with a stretch of at least 5 unpredictable amino acids as well as invariant amino acid sequences are included within the scope of the random peptides. The syngenes encoding such peptides will have both codons of unpredictable and codons of invariant sequence or (due to the degeneracy of the genetic code) degenerate codons thereof. However, the binding domains of syngenes are not cDNA or generated so as to be genomic sequences. The syngenes of the invention

are not sequences generated by a method comprising mutagenesis (even random mutagenesis) of a cDNA sequence or portion thereof or genomic sequence or portion thereof coding for a peptide having a  
5 predetermined activity.

By the methods of the present invention, the binding domains encoded by syngenes are advantageously identified from random peptide libraries. Typically, random peptide libraries will be encoded by synthetic  
10 oligonucleotides with at least 15 contiguous variant nucleotide positions having the potential to encode all 20 naturally occurring amino acids. The sequence of amino acids encoded by the variant nucleotides is unpredictable and substantially random. The terms  
15 "unpredicted", "unpredictable" and "substantially random" are used interchangeably with respect to the amino acids encoded and are intended to mean that the variant nucleotides at any given position encoding the binding domain of the syngene product are such that it  
20 cannot be predicted which of the 20 naturally occurring amino acids will appear at that position. These variant nucleotides are the product of random chemical synthesis. As will become clear, the biological random peptide libraries envisioned for use  
25 include those in which a bias has been introduced into the random sequence, e.g., to disfavor stop codon usage.

In a specific embodiment, a syngene of the invention encodes a peptide comprising a binding  
30 domain (which binds to a ligand of choice), in which the nucleotide sequence encoding the binding domain is a sequence identified by a method comprising screening a library of recombinant vectors, said vectors comprising unpredictable nucleotides arranged in one  
35 or more contiguous sequences, wherein the total number of unpredictable nucleotides is greater than or equal to 15 and less than or equal to about 300. In other



specific embodiments, the total number of such unpredictable nucleotides is in the range of 15-600, 24-600, 60-600, 15-120, 60-120, or 60-300.

5        5.1.1. CHEMICALLY SYNTHESIZED PEPTIDE LIBRARIES

          The peptide libraries used in the present invention may be libraries that are chemically synthesized *in vitro*. Examples of such libraries are given in Fodor, S., et al., 1991, Science 251:767-773, 10 which describes the synthesis of a known array of short peptides on an individual microscopic slide; Houghten, R., et al., 1991, Nature 354:84-86, which describes mixtures of free hexapeptides in which the first and second residues in each peptide were 15 individually and specifically defined; Lam, K., et al., 1991, Nature 354:82-84, which describes a split synthesis scheme; and Medynski, 1994, Bio/Technology 12:709-710, who describes split synthesis and T-bag synthesis methods as well. See also Gallop et al., 20 1994, J. Medicinal Chemistry 37(9):1233-1251.

          Screening of chemically synthesized peptide libraries to identify peptides which bind to a ligand of choice can be carried out by methods well known in the art.

25        In a specific embodiment, the total number of unpredictable amino acids in the peptides of the library used for screening is greater than or equal to 5 and less than or equal to 25; in other embodiments the total is in the range of 5-15 or 5-10 amino acids, 30 preferably contiguous amino acids.

          While a binding domain can be identified from chemically synthesized peptide libraries and an appropriate syngene synthesized to encode such a binding domain, such domains would be small (i.e. less 35 than 10 amino acids, and most probably 5-6 amino acids, in length). Therefore, this approach is less preferred than the biological peptide libraries

containing unpredictable sequences of greater length, described below.

#### 5.1.2. BIOLOGICAL PEPTIDE LIBRARIES

5 In another embodiment, biological random peptide libraries are used to identify a binding domain which binds to a ligand of choice. Many suitable biological random peptide libraries are known in the art and can be used to screen for a peptide  
10 that binds to a ligand of choice, according to standard methods commonly known in the art.

According to this second approach, involving recombinant DNA techniques, peptides have been expressed in biological systems as either soluble  
15 fusion proteins or viral capsid fusion proteins.

A number of peptide libraries according to this approach have used the M13 phage. Although the N-terminus of the viral capsid protein, protein III (pIII), has been shown to be necessary for viral  
20 infection, the extreme N-terminus of the mature protein does tolerate alterations such as insertions. Accordingly, various random peptide libraries, in which the diverse peptides are expressed as pIII fusion proteins, are known in the art; these libraries  
25 can be used to identify syngene-encoded binding domains by screening against a ligand of choice. Examples of such libraries are described below.

Scott and Smith, 1990, Science 249:386-390 describe construction and expression of an "epitope  
30 library" of hexapeptides on the surface of M13. Cwirla et al., 1990, Proc. Natl. Acad. Sci. USA 87: 6378-6382 also described a somewhat similar library of hexapeptides expressed as pIII gene fusions of M13 fd phage. PCT publication WO 91/19818 dated December 26,  
35 1991 by Dower and Cwirla describes a similar library of pentameric to octameric random amino acid sequences. Devlin et al., 1990, Science, 249:404-406,

described a peptide library of about 15 residues generated using an (NNS) coding scheme for oligonucleotide synthesis in which S is G or C. Christian et al., 1992, J. Mol. Biol. 227:711-718

5 described a phage display library expressing decapeptides. Lenstra, 1992, J. Immunol. Meth. 152:149-157 described construction of a library by a laborious process encompassing annealing oligonucleotides of about 17 or 23 degenerate bases

10 with an 8 nucleotide long palindromic sequence at their 3' ends.

Other biological peptide libraries which can be used include those described in U.S. Patent No. 5,270,170, dated December 14, 1993; and PCT

15 Publication No. WO 91/19818, dated December 26, 1991.

In a specific embodiment, the R8C random peptide library (described in Section 6.10 and Figures 11 and 12) is used. In another specific embodiment, the R26 peptide library (described in Section 6.9.4

20 and Figure 9) is used. In yet another specific embodiment, the DC43 peptide library (described in Section 6.9.5 and Figure 13 is used).

The protein pVIII is a major M13 viral capsid protein which can also serve as a site for

25 expressing peptides on the surface of M13 viral particles in the construction of random peptide libraries.

While it would be understood by one skilled in the art that as few as 5 amino acids can constitute

30 a binding domain, the average functional domain within a natural protein is considered to be about 40 amino acids. Thus, the random peptide libraries from which the binding domains encoded by the syngenes of the present invention are preferably identified encode

35 peptides having in the range of 5 to 200 total variant amino acids. Although it is contemplated that biologically expressed random peptide libraries

displaying short random inserts (i.e. less than 20 amino acids in length) could be used to identify syngenes of the invention, the most preferred binding domains will be identified from biologically expressed random peptide libraries in which the displayed peptide has 20 or greater unpredictable amino acids i.e. preferably in the range of 20 to 100, and most preferably 20 to 50 amino acids, as exemplified by the TSAR libraries described herein.

One of the objects of the present invention is to provide syngenes encoding binding domains of greater binding specificity than found in nature. To accomplish this, the invention preferably uses libraries of greater complexity than are commonly employed in the art. The conventional teaching in the random peptide library art is that the length of inserted oligonucleotides should be kept short, encoding preferably fewer than 15 and most preferably about 6-8 amino acids. However, not only can libraries encoding more than about 20 amino acids be constructed, but such libraries can be advantageously screened to identify peptides having binding specificity for a variety of ligands. Such libraries with longer length inserts are exemplified by the TSAR libraries, described in detail hereinbelow in Section 5.1.3 and its subsections.

Libraries composed of longer length oligonucleotides afford the ability to identify peptides in which a short sequence of amino acids is common to or shared by a number of peptides binding a given ligand, i.e., library members having shared binding motifs. The use of longer length libraries also affords the ability to identify peptides which do not have any shared sequences with other peptides but which nevertheless have binding specificity for the same ligand.

Libraries having large inserted oligonucleotide sequences provide the opportunity to identify or map binding sites which encompass not only a few contiguous amino acid residues, i.e., simple binding sites, but also those which encompass discontinuous amino acids, i.e., complex binding sites.

Additionally, the large size of the inserted synthesized oligonucleotides of certain libraries provides the opportunity for the development of secondary and/or tertiary structure in the potential binding peptides and in sequences flanking the actual binding site in the binding domain. Secondary and tertiary structure often significantly affect the ability of a sequence to mediate binding, as well as the strength and specificity of any binding which occurs. Such complex structural developments are not feasible when only small length oligonucleotides are used.

Finally, as has been overlooked by the conventional wisdom, longer length peptide libraries provide a greatly enhanced complexity over shorter length peptide libraries. This greatly enhanced complexity is associated with the concept of sliding windows which must be counted inclusively, i.e.,  $\text{number of windows} = [\text{length of sequence}] - [\text{window size}] + 1$ . This concept can be illustrated by comparison of two libraries, as follows. Assume that a binding site to a ligand requires 5 contiguous amino acid residues (pentamer). In two libraries composed of equal numbers of recombinants, a first library expressing pentamers, and another library constructed according to the present invention expressing 30-mers, the second library will be 26 times "richer" in binding sites relative to the first library, with the additional potential complexity contributed by

secondary or tertiary structure provided by flanking regions.

Th refore, it is contemplated that the most preferred binding domains and the syngenes encoding them will be identified from biologically expressed random peptide libraries in which the displayed peptide is 20 or greater amino acids in length. Examples of such random peptide libraries are the TSAR libraries described in detail in Section 5.1.3 and its subsections.

The most preferred libraries for practicing the invention are those that are generated or constructed to express a plurality of heterofunctional fusion peptides. The putative binding domains in these peptides are not known to be naturally occurring amino acid sequences or encoded by naturally occurring nucleotide sequences. Thus, for example, the random peptide libraries from which the binding domains encoded by syngenes are identified are not cDNA or genomic libraries. The sequence of any given peptide from the preferred libraries cannot be predicted in advance. In a preferred embodiment, the peptides are expressed on the surface of the recombinant vectors of the library.

In one embodiment, the random library is a linear, non-constrained library. As would be understood by one in the art having considered the present disclosure, in another specific embodiment, "constrained", "structured" or "semi-rigid" random peptide libraries could also be used in the present methods to identify binding domains encoded by syngenes. Typically, these libraries express peptides that are substantially random but contain a small percentage of fixed residues within or flanking the random sequences that have the result of conferring structure or some degree of conformational rigidity to the peptide. In a semirigid peptide library, the

plurality of synthetic oligonucleotides express peptides that are each able to adopt only one or a small number of different conformations that are constrained by the positioning of codons encoding certain structure conferring amino acids in or flanking the synthesized variant or unpredicted oligonucleotides. Unlike linear, unconstrained libraries in which the plurality of proteins expressed potentially adopt thousands of short-lived different conformations, in a semirigid peptide library, the plurality of proteins expressed can adopt only a single or a small number of conformations. Such libraries are exemplified by the TSAR-13 and TSAR-14 libraries described in Section 6.4 and its subsections; by a library of random 6 amino acid sequences, each flanked by invariant cysteine residues (O'Neil et al., 1992, Proteins 14:509-515); by a library of random 8 amino acid sequences, each flanked by invariant cysteine residues as in the R8C library described in Section 6.10 and Figures 11 and 12; and by those libraries disclosed in PCT Publication No. WO94/11496, dated May 26, 1994.

The DNA encoding the random peptides of the libraries can be synthetic or natural in origin. For example, DNA can be prepared by a method similar to methods used for constructing a cDNA library. However, the DNA is then sheared and/or digested into DNA pieces of 30 base pairs or smaller. The DNA pieces are then ligated together, preferably to sizes of about 120 base pairs to create DNA sequences that do not occur in nature. Peptides encoded by such sequences can be expressed in a library and screened for binding domains. However, as would be appreciated by one of skill in the art, such DNA sequences will contain stop codons and therefore would be a less desirable source of coding sequences. It is noted that such a library, by virtue of the juxtaposition of

cDNA sequences used in creating it, is not a cDNA library.

It is not intended by the present invention to use cDNA libraries for identifying syngenes. Such  
5 libraries have been screened with oligonucleotides (Staudt et al., 1988, Science 241:577-580; Singh et al., 1988, Cell 52:415-423). These approaches yielded naturally occurring DNA binding proteins. The object of syngenes is to identify genes encoding sequences  
10 that are not present in nature.

### 5.1.3. TSAR LIBRARIES

In a preferred embodiment, a biological peptide library that is a random peptide "TSAR"  
15 library is screened to identify the synthetic gene sequences encoding a binding domain, for use in constructing a syngene. In this embodiment, the syngenes of the present invention encode peptides called TSARs which bind to a ligand of choice. TSARs  
20 is an acronym for "Totally Synthetic Affinity Reagents" as described in PCT publication WO 91/12328, dated August 22, 1991. TSAR libraries, their construction and use, and specific examples of TSAR libraries are described in these publications and in  
25 detail below. As described herein, nucleic acids encoding TSARs or a TSAR portion which mediates binding to the ligand of choice can be used to identify and construct the syngenes of the present invention.

30 As used in the present invention, a TSAR is intended to encompass a concatenated heterofunctional peptide that includes at least two distinct functional regions. One region of the heterofunctional TSAR molecule is a binding domain with affinity for a  
35 ligand, that is preferably characterized by 1) its strength of binding under specific conditions, 2) the stability of its binding under specific conditions,



and 3) its selective specificity for the chosen ligand. A second region of the heterofunctional TSAR molecule is an effector domain which in specific embodiments of the TSAR libraries is the pIII protein or other structural protein providing for phage display of the heterofunctional peptide. In other embodiments, such a sequence can also include a sequence that is biologically or chemically active to enhance expression and/or detection and/or purification of the TSAR. Such a sequence can be chosen from a number of biologically or chemically active proteins including a structural protein or fragment that is accessibly expressed as a surface protein of a vector, an enzyme or fragment thereof, a toxin or fragment thereof, a therapeutic protein or peptide, or a protein or peptide whose function is to provide a site for attachment of a substance such as a metal ion, etc., that is useful for enhancing expression and/or detection and/or purification of the expressed TSAR.

A TSAR can contain an optional additional linker domain or region between the binding domain and the effector domain. The linker region serves (1) as a structural spacer region between the binding and effector domains; (2) as an aid to uncouple or separate the binding and effector domains; or (3) as a structural aid for display of the binding domain and/or the TSAR by the expression vector.

A TSAR may be a heterofunctional fusion protein, said fusion protein comprising (a) a binding domain encoded by an oligonucleotide comprising unpredictable nucleotides in which the unpredictable nucleotides are arranged in one or more contiguous sequences, wherein the total number of unpredictable nucleotides is greater than or equal to about 60 and less than or equal to about 600, and optionally, (b) an effector domain encoded by an oligonucleotide

sequence which is a protein or peptide that enhances expression or detection of the binding domain.

Alternatively, a TSAR may be a heterofunctional fusion protein comprising (i) a binding domain encoded  
5 by a double stranded oligonucleotide comprising unpredictable nucleotides in which the unpredictable nucleotides are arranged in one or more contiguous sequences, wherein the total number of unpredictable  
10 nucleotides is greater than or equal to about 60 and less than or equal to about 600 and the contiguous sequences are flanked by invariant residues designed to encode amino acids that confer a desired structure to the binding domain of the expressed  
heterofunctional fusion protein, and, optionally, (ii)  
15 an effector domain encoded by an oligonucleotide sequence encoding a protein or peptide that enhances expression or detection of the binding domain.

#### 5.1.3.1. CONSTRUCTION OF TSAR LIBRARIES

20 In one embodiment of the present invention, in order to identify and obtain the syngenes of the present invention, use is made of TSAR libraries. In order to prepare a library of vectors expressing a plurality of protein TSARs according to one embodiment  
25 of the present invention, single stranded sets of nucleotides are synthesized and assembled in vitro, by way of example, according to the following scheme.

The synthesized nucleotide sequences are designed to have variant or unpredicted as well as  
30 invariant nucleotide positions. Pairs of variant nucleotides in which one individual member is represented by 5' (NNB)<sub>n</sub> 3' and the other member is represented by 3' (NNV)<sub>m</sub> 5' where N is A, C, G or T; B is G, T or C; V is G, A or C; n is an integer, such  
35 that 10 ≤ n ≤ 100, and m is an integer, such that 10 ≤ m ≤ 100, are synthesized for assembly into synthetic oligonucleotides. As assembled, there are at least n

+ m variant codons in each inserted synthesized double stranded oligonucleotide sequence.

As would be understood by those of skill in the art, the variant nucleotide positions have the potential to encode all 20 naturally occurring amino acids and, when assembled as taught by the present method, encode only one stop codon, i.e., TAG. The sequence of amino acids encoded by the variant nucleotides is unpredictable and substantially random in sequence. As explained hereinabove, the terms "unpredicted", "unpredictable" and "substantially random" are used interchangeably in the present application with respect to the amino acids encoded and are intended to mean that at any given position within the binding domain of the TSARs encoded by the variant nucleotides which of the 20 naturally occurring amino acids will occur cannot be predicted.

The variant nucleotides, according to the TSAR scheme, encode all twenty naturally occurring amino acids by use of 48 different codons.

Invariant nucleotides are positioned at particular sites in the nucleotide sequences to aid in assembly and cloning of the synthesized oligonucleotides. At the 5' termini of the sets of variant nucleotides, the invariant nucleotides encode for efficient restriction enzyme cleavage sites. The invariant nucleotides at the 5' termini are chosen to encode pairs of sites for cleavage by restriction enzymes (1) which can function in the same buffer conditions; (2) are commercially available at high specific activity; (3) are not complementary to each other to prevent self-ligation of the synthesized double stranded oligonucleotides; and (4) which require either 6 or 8 nucleotides for a cleavage recognition site in order to lower the frequency of cleaving within the inserted double stranded synthesized oligonucleotide sequences. According to

one particular method of constructing peptide libraries, the selected restriction site pairs are selected from Xho I and Xba I, and Sal I and Spe I. Other examples of useful restriction enzyme sites include, but are not limited to: Nco I, Nsi I, Pal I, Not I, Sfi I, Pme I, etc. Restriction sites at the 5' termini invariant positions function to promote proper orientation and efficient production of recombinant molecule formation during ligation when the oligonucleotides are inserted into an appropriate expression vector.

According to an alternative method of constructing peptide libraries, the variant nucleotides are synthesized using one or more methylated dNTPs and the 5' termini invariant nucleotides, encoding restriction sites for efficient cleavage, are synthesized using non-methylated dNTPs. This embodiment provides for efficient cleavage of long length synthesized oligonucleotides at the termini for insertion into an appropriate vector, while avoiding cleavage in the variant nucleotide sequences.

The 3' termini invariant nucleotide positions are complementary pairs of 6, 9 or 12 nucleotides to aid in annealing two synthesized single stranded sets of nucleotides together and conversion to double-stranded DNA, designated herein synthesized double stranded oligonucleotides.

In particular peptide libraries, the 3' termini invariant nucleotides are selected from 'GCGGTG' and 'CGCCAC', and 'CCAGGT' and 'GGTCCA', which also encode either a particular amino acid, glycine, or dipeptide proline-glycine, which provides the flexibility of either a swivel or hinge type configuration to the expressed proteins, polypeptides and/or peptides, respectively.

In yet another specific library, the 3' termini invariant nucleotides of the coding strand are 5' GGGTGC GGC 3' which encode glycine, cysteine, glycine. In an oxidizing environment the cysteine  
5 forms a disulfide bond with another cysteine engineered into the binding domain to form a semirigid conformation in the expressed peptide.

In another library, the complementary 3' termini also encode an amino acid sequence that  
10 provides a short charge cluster (for example, KKKK (SEQ ID NO: 49), DDDD (SEQ ID NO: 50) or KDKD (SEQ ID NO: 51)), or a sharp turn (for example, NPXY (SEQ ID NO: 52), YXRF (SEQ ID NO: 53) where X is any amino acid). In another alternative library, the  
15 complementary 3' termini also encode a short amino acid sequence that provides a peptide known to have a desirable binding or other biological activity. Specific examples include complementary pairs of sequences encoding peptides including but not limited  
20 to RGD, HAV, HPQ $\theta$  where  $\theta$  is a non-polar amino acid. These short amino acid sequences are intended to aid in screening the library and retrieving members of the library; they are not intended to be binding domains.

Figure 1A generally illustrates an assembly  
25 process. The oligonucleotide sequences are thus assembled by a process comprising: synthesis of pairs of single stranded nucleotides having a formula represented by:

(a) 5'  $\rightarrow$  3' Restriction site-(NNB)<sub>n</sub>-Complementary site;  
30 and

(b) 3'  $\rightarrow$  5' Complementary site-(NNV)<sub>m</sub>-Restriction site,  
where n is an integer, such that  $10 \leq n \leq 100$  and m is an integer, such that  $10 \leq m \leq 100$ . More particularly,  
the single stranded nucleotides are represented as:  
35 pairs of nucleotide sequences of a first formula

5' X (NNB)<sub>n</sub> J Z 3' and a second nucleotide sequence of the formula 3' Z' O U (NNV)<sub>m</sub> Y 5' where X and Y are restriction enzyme recognition sites, such that X ≠ Y;

N is A, C, G or T;

5 B is G, T or C;

V is G, A or C;

n is an integer, such that 10 ≤ n ≤ 100;

m is an integer, such that 10 ≤ m ≤ 100;

Z and Z' are each a sequence of 6, 9 or 12

10 nucleotides, such that

Z and Z' are complementary to each other; and

J is A, C, G, T or nothing;

O is A, C, G, T or nothing; and

U is G, A, C or nothing; provided, however, if

15 any one of J, O or U is nothing then J, O and U are all nothing.

Any method for synthesis of the single stranded sets of nucleotides is suitable, including use of an automatic nucleotide synthesizer. The synthesizer can be programmed so that the nucleotides can be incorporated, either in equimolar or non-equimolar ratios at the variant positions, i.e., N, B, V, J, O or U. The nucleotide sequences of the desired length are purified, for example, by HPLC.

25 Pairs of the purified, single stranded nucleotides of the desired length are reacted together in appropriate buffers through repetitive cycles of annealing and DNA synthesis using an appropriate DNA polymerase, such as Taq, Vent™ or Bst DNA polymerase, and appropriate temperature cycling. Sequenase™ (modified T7 DNA polymerase) is preferred for use, and can be employed according to the instructions of the manufacturer (U.S. Biochemical, Cleveland, OH), without temperature cycling. Klenow fragment of E. coli DNA polymerase could be used but, as would be understood by those of skill in the art, such polymerase would need to be replenished at each cycle

and thus is less preferred. The double stranded DNA reaction products, now greater than  $m + n$  in length, are isolated, for example, by phenol/chloroform extraction and precipitation with ethanol.

5           After resuspension in buffer, the double stranded synthetic oligonucleotides are cleaved with appropriate restriction enzymes to yield a plurality of synthesized oligonucleotides. The double-stranded synthesized oligonucleotides should be selected for  
10 those of the appropriate size by means of high resolution polyacrylamide gel electrophoresis, or NuSieve/MetaMorph (FMC Corp., Rockland, MA) agarose gel electrophoresis, or the like. Size selection of the oligonucleotides substantially eliminates abortive  
15 assembly products of inappropriate size and incomplete digestion products.

          This scheme for synthesis and assembly of the unpredictable oligonucleotides used to construct the TSAR libraries incorporates  $m + n$  variant,  
20 unpredicted nucleotide sequences of the formula  $(NNB)_{n,m}$  where B is G, T or C and  $n$  and  $m$  are each an integer, such that  $20 \leq n + m \leq 200$ ; thus, from 20 to 200 unpredicted codons are incorporated into the synthesized double stranded oligonucleotides. Such a  
25 scheme provides a number of important advantages not available with conventional libraries. As assembled, the synthesized oligonucleotides encode all twenty naturally occurring amino acids by use of 48 different amino acid encoding codons. Although this uses  
30 somewhat less variability than that found in nature where 64 different codons are used, the scheme advantageously provides greater variability than other conventional schemes. For example, conventional schemes in which the variant nucleotides have the  
35 formula  $NNK$ , where K is G or T (see Dower, PCT Publication No. WO 91/19818), or the formula  $NNS$ , where S is C or G (see Devlin, PCT Publication No.

WO 91/18980), use only 32 different amino acid-encoding codons. The use of a larger number of amino acid encoding codons may make the TSAR libraries less susceptible to codon preferences of the host when the libraries are expressed. Although both the TSAR scheme and conventional schemes retain only 1 stop codon, use of NNB as taught in the TSAR scheme advantageously provides synthesized oligonucleotides in which the probability of a stop codon is decreased compared to conventional NNS or NNK schemes.

Perhaps most significantly, the TSAR scheme for synthesis and assembly of the oligonucleotides provides sequences of oligonucleotides encoding unpredicted amino acid sequences which are larger in size than conventional libraries. The present synthesized double stranded oligonucleotides comprise at least about 77-631 nucleotides encoding the restriction enzyme sites, the complementary site and about 20-200 unpredicted amino acids in the TSAR binding domain. According to a preferred embodiment,  $n$  and  $m$  are greater than or equal to 10 and less than or equal to 50. Thus, the synthesized double stranded oligonucleotides comprise at least 77-331 nucleotides and encode about 20-100 unpredicted amino acids in the TSAR binding domain. In specific examples, the synthesized oligonucleotides encode 20, 24 and 36 unpredicted amino acids and 27, 35 and 42 total amino acids, respectively for the TSAR-9, TSAR-12, and TSAR-13 libraries, in the TSAR binding domain.

According to an alternative embodiment, syngenes are isolated from a library which expresses a plurality of TSAR peptides having some degree of conformational rigidity in their structure (constrained or semirigid peptide libraries, illustrated in Figure 1D-F and exemplified by the TSAR-13 and TSAR-14 libraries described in Section 6.4 and its subsections.



At least four different methods can be used to engineer TSAR libraries used to identify syngene sequences encoding a binding domain, so that the expressed peptides are semirigid or have some degree of conformational rigidity. In the first method, the synthesized oligonucleotides are designed so that the expressed peptides have a pair of invariant cysteine residues positioned in, or flanking, the unpredicted or variant residues (See Figure 1D). When the library is expressed in an oxidizing environment, the cysteine residues should be in the oxidized state, most likely cross-linked by disulfide bonds to form cystines. Thus, the peptides would form rigid or semirigid loops. The nucleotides encoding the cysteine residues should be placed from 6 to 27 amino acids apart flanking the variant nucleotide sequences. A particular peptide library having such structure is the R8C library described in Section 6.13 *infra*.

The actual positions of the invariant residues can be modeled on the arrangement observed in isolates from a linear peptide library, for example, TSAR peptides in which two or four cysteines are encoded by the inserted synthesized oligonucleotides, isolated from the TSAR-9 or TSAR-12 libraries (see Section 6.9 and its subsections *infra*). The following general formulas illustrate the structure of these peptides:

- (1)  $X(NNB)_6(TGC)(NNB)_{11}Z(NNB)_{14}(TGC)(NNB)_3Y$
- (2)  $X(NNB)_1(TGC)(NNB)_{10}(TGC)_2(NNB)_4Z(NNB)_6(TGC)(NNB)_3Y;$
- (3)  $X(NNB)_{16}(TGC)(NNB)_1Z(NNB)_{16}(TGC)(NNB)_1Y;$
- (4)  $X(NNB)_{11}(TGC)(NNB)_6Z(NNB)_7(TGC)(NNB)_{10}Y$

The positions of the cysteines are well tolerated as these phage are stable and infectious.

In the second method, a double stranded oligonucleotide sequence providing a cloverleaf

structure (see Figure 1E) can be represented, for example, by the formula:

$X(TGC)_1(NNB)_{10}(TGC)_1(NNB)_2Z(NNB)_2(TGC)_1(NNB)_{14}(TGC)_1Y$ .

When these peptides are expressed by the appropriate  
5 vectors, the cysteine residues may adopt three different disulfide bond arrangements, thereby generating three different patterns of "cloverleaves". The plurality of proteins, polypeptides and/or peptides expressed by this type of rigid library  
10 should form many different ligand binding pockets from which to select the best fit. It should be noted that when a semirigid library of the first or second type above is expressed in a viral vector in an oxidizing environment, there will likely be a selection against  
15 odd numbers of cysteines occurring within the unpredicted or random peptide regions expressed because one unpaired cysteine residue will likely cross-link the viral vectors and make them non-infectious. Examples of semi-rigid libraries  
20 providing a cloverleaf structure are the TSAR-13 and TSAR-14 libraries described in Sections 6.9.4 and 6.9.4.1 *infra*.

In the third method, the synthesized nucleotides are designed and assembled so that the  
25 plurality of proteins expressed have both invariant cysteine and histidine residues positioned within the variant nucleotide sequences (see Figure 1F). The positions of the invariant residues can be modeled after the arrangement of cysteine and histidine  
30 residues seen in zinc finger proteins (i.e.,  $-CX_2, CX_2, HX_2, H-$ , where X is any amino acid), thereby creating a library of zinc finger-like proteins. As used herein the term "zinc finger-like proteins" is intended to mean any of the plurality of proteins  
35 expressed which contain invariant cysteine and histidine residues which confer a zinc finger or similar structure on the expressed protein.

In the fourth method, (see Figure 1F), the plurality of proteins are designed to have invariant histidine residues positioned within the variant nucleotide sequences. To illustrate, exemplary histidine containing TSARs can be represented by the following general formulas:

- (1)  $X(NNB)_a(CAC)(NNB)_b(CAC)(NNB)_cZ(NNB)_d(CAC)(NNB)_e(CAC)_f(NNB)_gY$ ;
- (2)  $X(NNB)_a(CAC)(NNB)_b(CAC)(NNB)_cZ(CAC)(NNB)_d(CAC)_e(NNB)_f(CAC)(NNB)_gY$ ;
- (3)  $X(NNB)_a(CAC)(NNB)_b(CAC)_c(NNB)_d(CAC)(NNB)_eZ(NNB)_f(CAC)_g(NNB)_hY$ ; and
- (4)  $X(CAC)(NNB)_a(CAC)(NNB)_b(CAC)(NNB)_c(CAC)(NNB)_dZ(CAC)(NNB)_e(CAC)(NNB)_f(CAC)(NNB)_gY$ , where CAC represents the codon for histidine.

To maintain the rigid cloverleaf conformation of this plurality of proteins, the TSAR proteins are expressed and harvested in the presence of 1-1000  $\mu$ M zinc chloride. The expressed proteins could also be saturated with other divalent metal cations, such as  $Cu^{2+}$  and  $Ni^{2+}$ . The members of this type of rigid library may have advantageous chemical reactivity, since metal ions are often within the catalytic sites of enzymes.

To prepare a library of syngenes encoding semi-rigid or constrained peptides, the synthesized single stranded nucleotides are assembled by annealing a first nucleotide sequence of the formula:

5'X [ $\alpha$  (NNB) $_a$ ] $_c$  JZ 3' with a second nucleotide sequence of the formula

3' Z'OU [(NNV) $_b$ ] $_d$  Y 5'

where a, c, b, d are integers such that  $20 \leq [a]_c + [b]_d \leq 200$ ; and c and d are each  $\geq 1$ ;

$\alpha$  is an invariant nucleotide sequence that confers some structure in the peptide it encodes and  $\beta$  is an invariant nucleotide sequence whose complimentary nucleotide sequence confers some structure in the encoded peptide; and  
 5 X, Y, N, B, V, Z, Z', J, O, U are as defined above.

This scheme for synthesis of unpredictable oligonucleotides incorporates a total of the  
 10 arithmetic sum of  $(a \times c) + (b \times d)$ , i.e.,  $[a]_c + [b]_d$  variant, unpredicted nucleotide sequences, i.e.,  $[(NNB)_c]_c + [(NNV)_d]_d$ , flanked by invariant nucleotides, i.e.,  $\alpha$  and  $\beta$ , which encode structure-conferring amino acid sequences.

15 By way of example,  $\alpha$  and  $\beta$  could include a codon for one or more cysteine residues, for example Gly-Cys-Gly, in which instance a and b are each preferably  $\geq 6$  and  $\leq 27$ , to generate disulfide bonds between different cysteines in the expressed loop  
 20 forming peptide structures. Appropriate oligonucleotides could be assembled, by annealing for example, a first nucleotide sequence of the formula:  
 5' X  $\alpha$  (NNB)<sub>c</sub>  $\alpha$  (NNB)<sub>c</sub> Z 3' with a second nucleotide sequence of the formula

25 3' Z' (NNV)<sub>d</sub>  $\beta$  Y 5'.

More particularly, where  $\alpha$  encodes the sequence Gly-Cys-Gly (and the complementary sequence of  $\beta$  encodes the same sequence, and where both a and b are equal to seven, the synthesized single stranded nucleotides are  
 30 assembled by annealing a first nucleotide sequence:

5' X (GGG) (TGT) (GGG) (NNB)<sub>c</sub> (GGG) (TGT) (GGG) (NNB)<sub>c</sub> (GGG) (TGT) (GGG) 3' with a second nucleotide sequence:  
 3' (CCC) (ACA) (CCC) (NNV)<sub>d</sub> (CCC) (ACA) (CCC) Y 5'

where GGG represents the codon for glycine and TGT  
 35 represents the codon for cysteine. This oligonucleotide scheme encodes peptides, whose amino acid sequence would be GCGX<sub>c</sub>GCGX<sub>c</sub>GCGX<sub>c</sub>GCG.

Alternatively,  $\alpha$  and  $\beta$  could encode one or more histidine residues, for example GHGHH (SEQ ID NO: 54). In yet another alternative embodiment,  $\alpha$  and  $\beta$  could encode a Leu residue in which instance a and b are each  $\leq$  about 7. Such alternative embodiment would provide an alpha helical structure in the expressed peptides.

Additionally, according to yet another alternative embodiment, an  $\alpha$  group could be used for Z, and  $\beta$  for Z' to provide the complementary sequences to aid in annealing the nucleotides.

Other nucleotide sequences encoding amino acids that will impose structural constraints on the expressed peptides are possible as would become apparent to one of skill in the art based on the above description and syngenes encoding such constrained peptides are encompassed within the scope of the present invention.

An additional feature of these semirigid libraries is the potential to control the binding properties of isolates by reversibly destroying or altering the rigidity of the peptide. For example, it should be possible to elute a TSAR bound to a particular ligand in a gentle manner with reducing agents (i.e., DTT,  $\beta$ -mercaptoethanol) or divalent cation chelators (i.e., EDTA, EGTA). Such reagents can be used, for example, to elute a TSAR library expressed on phage vectors from target ligands. EDTA or EGTA, at low concentrations, does not appear to disrupt phage integrity or infectivity.

Once the phage have been recovered and it is deemed necessary to remove thiols from the solution, the reduced cysteine residues can be alkylated with iodoacetamide. This treatment prevents renewed disulfide bond formation and only diminishes phage infectivity 10-100 fold, which is tolerable since phage cultures usually attain titers of  $10^{12}$  plaque

forming units per milliliter. Alternatively, the elution reagents can be removed by dialysis (i.e., dialysis bag, Centricon/Amicon microconcentrators).

5

#### 5.1.3.2. EXPRESSION OF VECTORS ENCODING TSARS

In constructing a TSAR library, the plurality of oligonucleotides of appropriate size prepared as described above is inserted into an appropriate vector which, when inserted into a suitable host, expresses the plurality of peptides as heterofunctional fusion proteins with an expressed component of the vector (effector domain) which are screened to identify TSARs having affinity for a ligand of choice. According to an optional embodiment, the plurality of peptides further comprise a linking domain between the binding and effector domains. In a preferred mode of this embodiment, the linker domain is expressed as a fusion protein with the effector domain of the vector into which the plurality of oligonucleotides are inserted.

The skilled artisan will recognize that to achieve transcription and translation of the plurality of TSAR encoding oligonucleotides, the synthetic oligonucleotides must be placed under the control of a promoter compatible with the chosen vector-host system. A promoter is a region of DNA at which RNA polymerase attaches and initiates transcription. The promoter selected may be any one that has been synthesized or isolated that is functional in the vector-host system. For example, *E. coli*, a commonly used host system, has numerous promoters such as the *lac* or *trp* promoter or the promoters of its bacteriophages or its plasmids. Also synthetic or recombinantly produced promoters such as the  $p_{TAC}$  promoter may be used to direct high level expression of the gene segments adjacent to them.

Signals are also necessary in order to attain efficient translation of the inserted oligonucleotides. For example in *E. coli* mRNA, a ribosome binding site includes the translational start codon AUG or GUG in addition to other sequences complementary to the bases of the 3' end of 16S ribosomal RNA. Several of these latter sequences such as the Shine/Dalgarno (S/D) sequence have been identified in *E. coli* and other suitable host cell types. Any S/D-ATG sequence which is compatible with the host cell system can be employed. These S/D-ATG sequences include, but are not limited to, the S/D-ATG sequences of the *cro* gene or *N* gene of bacteriophage lambda, the tryptophan *E*, *D*, *C*, *B* or *A* genes, a synthetic S/D sequence or other S/D-ATG sequences known and used in the art. Thus, regulatory elements control the expression of the polypeptide or proteins to allow directed synthesis of the reagents in cells and to prevent constitutive synthesis of products which might be toxic to host cells and thereby interfere with cell growth.

Any of a variety of vectors can be used, including, but not limited to bacteriophage vectors such as  $\phi$ X174,  $\lambda$ , M13 and its derivatives, f1, fd, Pfl, etc., phagemid vectors, plasmid vectors, insect viruses, such as baculovirus vectors, mammalian cell vectors, including such virus vectors as parvovirus vectors, adenovirus vectors, vaccinia virus vectors, retrovirus vectors, etc., yeast vectors such as Ty1, killer particles, etc.

An appropriate vector contains or is engineered to contain a gene encoding an effector domain of a TSAR comprising the pIII protein or other suitable protein or portion thereof providing for phage display. The effector domain gene contains or is engineered to contain multiple cloning sites. At least two different restriction enzyme sites within

such gene, comprising a polylinker, are preferred. The vector DNA is cleaved within the polylinker using two different restriction enzymes to generate termini complementary to the termini of the double stranded synthesized oligonucleotides assembled as described above. Preferably the vector termini after cleavage have or are modified, using DNA polymerase, to have non-compatible sticky ends that do not self-ligate, thus favoring insertion of the double-stranded synthesized oligonucleotides and hence formation of recombinants expressing the TSAR fusion proteins, polypeptides and/or peptides. The double stranded synthesized oligonucleotides are ligated to the appropriately cleaved vector using DNA ligase.

It is particularly useful to include a "stuffer fragment" within the polylinker region of the vector when the vector (e.g. phage or plasmid) is intended to express the TSAR as a heterofunctional fusion protein that is expressed on the surface of the vector. As used in the present application, a "stuffer fragment" is intended to encompass a relatively short, i.e., about 24-45 nucleotides, known DNA sequence flanked by at least 2 restriction enzyme sites, useful for cloning, said DNA sequences coding for a binding site recognized by a known ligand, such as an epitope of a known monoclonal antibody. The restriction enzyme sites at the termini of the stuffer fragment are useful for insertion of the synthesized double stranded oligonucleotides, resulting in deletion of the stuffer fragment.

Because of the physical linkage between the expressed heterologous fusion protein and the phage or plasmid vector containing the stuffer fragment and because the stuffer fragment can comprise a known DNA sequence encoding a protein that is immunologically active (i.e., an immunological marker), the presence or absence of the stuffer fragment can be easily



detected either at the nucleotide level, by DNA sequencing, PCR or hybridization, or at the amino acid level, e.g., using an immunological assay. Such determination allows rapid discrimination between  
5 recombinant (TSAR expressing) vectors generated by insertion of the synthesized double stranded oligonucleotides and non-recombinant vectors.

According to a preferred embodiment, the stuffer fragment comprises the DNA fragment encoding  
10 the epitope of the human c-myc protein recognized by the murine monoclonal antibody 9E10 having the amino acid sequence EQKLISEEDLN (SEQ ID NO: 55) (Evan et al., 1985, Mol. Cell. Biol. 5:3610-3616) with a short flanking sequence of amino acids at the 5' and 3'  
15 termini which serve as restriction enzyme sites so that the stuffer fragment can be removed and the synthesized double stranded oligonucleotides can be inserted using the restriction sites.

In another aspect, the stuffer fragment  
20 provides an efficient means to remove any non-recombinant vectors to enhance or enrich the population of TSAR expressing vectors, if necessary. Because the stuffer fragment is expressed e.g., as an immunologically active surface protein on the surface  
25 of non-recombinant vectors, it provides an accessible target for binding e.g., to an immobilized antibody. The non-recombinants thus could be easily removed from a library for example by serial passage over a column having the antibody immobilized thereon to enrich the  
30 population of recombinant TSAR-expressing vectors in the library.

In a preferred embodiment, the vector providing for expression of the TSAR libraries is or is derived from a filamentous bacteriophage, including  
35 but not limited to M13, f1, fd, Pf1, etc. vector encoding a phage structural protein, preferably a phage coat protein, such as pIII, pVIII, etc. In a

more preferred embodiment, the filamentous phage is an M13-derived phage vector such as m655, m663, and m666 described in Fowlkes et al., 1992, BioTechniques, 13:422-427 which encodes the structural coat protein pIII.

The phage vector is chosen to contain or is constructed to contain a cloning site located in the 5' region of a gene encoding a bacteriophage structural protein so that the plurality of synthesized double stranded oligonucleotides inserted are expressed as fusion proteins on the surface of the bacteriophage. This advantageously provides not only a plurality of accessible expressed peptides but also provides a physical link between the peptides and the inserted oligonucleotides to provide for easy screening and sequencing of the identified TSARs. Alternatively, the vector is chosen to contain or is constructed to contain a cloning site near the 3' region of a gene encoding structural protein so that the plurality of expressed proteins constitute C-terminal fusion proteins.

According to a preferred embodiment, the structural bacteriophage protein is pIII. The m663 vector described by Fowlkes et al. (1992, BioTechniques 13:422-427), containing the pIII gene having a c-myc-epitope (comprising the "stuffer fragment") introduced at the N-terminal end, flanked by Xho I and Xba I restriction sites may be used. The library may be constructed by cloning the plurality of synthesized oligonucleotides into a cloning site near the N-terminus of the mature coat protein of the appropriate vector, preferably the pIII protein, so that the oligonucleotides are expressed as coat protein-fusion proteins.

Alternatively, the plurality of oligonucleotides is inserted into a phagemid vector. Phagemids are utilized in combination with a defective

helper phage to supply missing viral proteins and replicative functions. Helper phage useful for propagation of M13 derived phagemids as viral particles include but are not limited to M13 phage K07, R408, VCS, etc. Generally, according to a preferred mode of this embodiment, the appropriate phagemid vector is constructed by engineering the Bluescript II SK+ vector (GenBank #52328) (Alting-Mees et al., 1989, Nucl. Acid Res. 17:9494); to contain (1) a truncated portion of the M13 pIII gene, i.e., nucleotides encoding amino acid residues 198-406 of the mature pIII, (2) the PelB signal leading with an upstream ribosome binding site and a short polylinker of Pst I, Xho I, Hind III, and Xba I restriction sites, in which the Xho I and Xba I sites are positioned so the synthesized double stranded oligonucleotides could be cloned and expressed in the same reading frame as the m663 phage vector; and (3) the linker sequence encoding GGGGS (SEQ ID NO: 56) between the polylinker and the pIII gene.

Alternatively, the synthesized oligonucleotides are inserted into a plasmid vector. An illustrative suitable plasmid vector for expressing the TSAR libraries is a derivative of plasmid p340-1 (ATCC No. 40516).

In order to obtain the appropriate p340-1 derivative suitable as an expression vector, the Nco I - Bam HI fragment is removed from p340-1 plasmid and replaced by a double stranded sequence having Xho I and Xba I restriction sites in the correct reading frame. In practice, p340-1 is cleaved using restriction enzymes at the Bgl II and Xba I sites and annealed with two oligonucleotides:

(1) 5'-CATGGCTCGAGGCTGAGTTCTAGA-3' (SEQ ID NO: 57) and  
(2) 5'-GATCTCTAGAACTCAGCCTCGAGC-3' (SEQ ID NO: 58)

having Nco I and Bam HI sticky ends. After ligation and transformation of *E. coli*, recombinants containing

the desired plasmid designated p340-1D are selected based on the inserted SEQ ID NOS: 57 and 58 and verified by sequencing. Like the parent p340-1, the desired p340-1D does not produce functional  $\beta$ -galactosidase because this gene is out of frame. Thus, when the synthesized double stranded oligonucleotides are inserted, using the Xho I and Xba I restriction sites, into the p340-1D vector the coding frame is restored and the TSAR binding domain is expressed as a fusion protein with the  $\beta$ -galactosidase. When exposed to IPTG, the vectors expressing the TSAR library would produce identifiable blue colonies.

Another illustrative plasmid vector useful to express a TSAR library is a plasmid derivative of plasmid pTrc99A (Amann et al., 1988, Gene 69:301-315) (Pharmacia, Piscataway, NJ) designated plasmid pLamB which is constructed to contain the LamB protein gene of *E. coli* (Clement and Hofnung, 1981, Cell 27:507-514) having a cloning site so that the plurality of oligonucleotides inserted are expressed as fusion proteins of the LamB protein.

Once the appropriate expression vectors are prepared, they are inserted into an appropriate host, such as *E. coli*, *Bacillus subtilis*, insect cells, mammalian cells, yeast cells, etc., for example by electroporation, and the plurality of oligonucleotides is expressed by culturing the transfected host cells under appropriate culture conditions for colony or phage production. Preferably, the host cells are protease deficient, and may or may not carry suppressor tRNA genes.

A small aliquot of the electroporated cells are plated and the number of colonies or plaques are counted to determine the number of recombinants. The library of recombinant vectors in host cells is plated

at high density for a single amplification of the recombinant vectors.

For example, recombinant M13 vector m666, m655 or m663, engineered to contain the synthesized double stranded oligonucleotides, is transfected into DH5 $\alpha$ F' *E. coli* cells by electroporation. TSARs are expressed on the outer surface of the viral capsid extruded from the host *E. coli* cells and are accessible for screening. The parent m666, m655 or m663 vectors contain the c-myc epitope (stuffer fragment). When the double stranded synthesized oligonucleotides are inserted between the Xho I and Xba I sites, the stuffer fragment is removed. The cloning efficiency of the expressed library is easily determined by filter blotting with the 9E10 antibody that recognizes the c-myc epitope.

Alternatively, when the double stranded synthesized oligonucleotides are cloned just at the Xho I or Xba I site, the c-myc epitope is retained. Then the c-myc epitope is expressed in the pIII-fusion protein expressed by the vector. An advantage of the m663 vector is that it contains an intact *LacZ'* gene, which can be easily seen as a blue dot when expressed in *E. coli* plated on X-gal and IPTG.

TSARs can be expressed in a plasmid vector contained in bacterial host cells such as *E. coli*. The TSAR proteins accumulate inside the *E. coli* cells and a cell lysate is prepared for screening. Use of plasmid p340-1D is described as an illustrative example. A TSAR library in p340-1D as described above, expressed the co-functional fusion protein with  $\beta$ -galactosidase. In the parent vector (without synthetic oligonucleotide) the  $\beta$ -galactosidase gene is out of frame and therefore nonfunctional. When plated on LB plates with ampicillin, IPTG and Xgal, the colonies that have TSAR oligonucleotides yield blue colonies, whereas colonies harboring non-recombinant

p340-1D or p340-1D r combinants with oligonucleotides carrying unsuppressed stop codons will be white. The relative number of blue and white colonies reveals the percent recombinants, and is useful in estimating the total numbers of recombinants in the library, and is also useful in screening.

Phagemid vectors containing the synthesized double stranded oligonucleotides, expressed on the outer surface of the extruded phage, are propagated either as infected bacteria or as bacteriophage with helper phage.

The expressed pDAF2-3 phagemids (See Section 6.10 and its subsections) have the added advantage that they include the c-myc gene which can serve as an "epitope tag" for the fusion pIII proteins. Approximately 0.1-10% of the phage carrying the phagemid genome incorporate the fusion pIII molecule. The intactness of the chimeric pIII proteins is evaluated based on the expression of the c-myc epitope. By following the expression of the c-myc epitope using the 9E10 antibody, it is possible to monitor the successful incorporation of the fusion pIII molecule into the M13 viral particle.

Also when expressing pDAF2, if the upstream c-myc peptide is detected immunologically using the 9E10 antibody, then it can be assumed that the downstream synthesized oligonucleotide, expressed TSAR peptide is appropriately expressed.

In addition, it may be of value to electroporate several different strains of *E. coli* and establish different versions of the same library. Of course, the same *E. coli* strain would need to be used for the entire set of screening experiments. This strategy is based on the consideration that there is likely an *in vivo* biological selection, both positive and negative, on the viral assembly, secretion, and infectivity rate of individual M13 recombinants due to

the sequence nature of the peptide-pIII fusion proteins. Therefore, *E. coli* with different genotypes (i.e., chaperonin overexpressing, or secretion enhanced) will serv as bacterial hosts, because they  
5 will yield libraries that differ in subtle, unpredictable ways.

5.2. SCREENING OF PEPTIDE LIBRARIES  
TO IDENTIFY BINDING DOMAINS  
ENCODED BY SYNGENES

10 The desired random peptide library is screened to identify and recover a syngene encoding a binding domain that binds to a ligand of choice.

As used in the present invention, a ligand  
15 is a substance for which it is desired to isolate a specific binding partner from a synthetic random peptide library. The term "ligand" is thus intended to include but not be limited to a substance, including a molecule or portion thereof, for which a  
20 proteinaceous receptor naturally exists or can be prepared according to the method of the invention. For example, a binding domain which binds to a ligand can function as a receptor, i.e., a lock into which the ligand fits and binds; or a binding domain can function as a key which fits into and binds a ligand  
25 when the ligand is a larger protein molecule. In this invention, a ligand includes, but is not limited to, a non-ionic chemical group, an organic chemical group, an ion, a metal, a metal or non-metal inorganic ion, a glycoprotein, a protein, a polypeptide, a peptide, a  
30 nucleic acid, a carbohydrate or carbohydrate polymer, a lipid, a fatty acid, a viral particle, a membrane vesicle, a cell wall component, a synthetic organic compound, a bioorganic compound and an inorganic compound or any portion of any of the above. Ligands  
35 also include the variable region of an antibody, an enzyme/substrate binding site, an enzyme/co-factor

binding site, a regulatory DNA binding protein, an RNA binding protein, a binding site of a metal binding protein, a nucleotid fold or GTP binding protein, a calcium binding protein, a membrane protein, a viral protein and an integrin. In another embodiment, the ligand is a peptide that is an intracellular targeting or processing signal (e.g., nuclear localization signals), e.g., whereby the syngene-encoded binding peptide which recognizes it will interfere with proper targeting of *in vivo* proteins containing such a targeting signal.

A preferred method for identifying syngenes that encode a binding domain that binds to a ligand of choice comprises screening a library of recombinant vectors that express a plurality of heterofunctional fusion proteins, said fusion proteins comprising (a) a binding domain encoded by an oligonucleotide comprising unpredictable nucleotides in which the unpredictable nucleotides are arranged in one or more contiguous sequences, wherein the total number of unpredictable nucleotides is greater than or equal to about 60 and less than or equal to about 600, and optionally, (b) an effector domain encoded by an oligonucleotide sequence which is a protein or peptide that enhances expression or detection of the binding domain. Screening is done by contacting the plurality of heterofunctional fusion proteins with the ligand of choice under conditions conducive to ligand binding and then isolating the fusion proteins which bind to the ligand. The methods of the invention further preferably comprise determining the nucleotide sequence encoding the binding domain of the heterofunctional fusion protein identified to determine the syngene sequence that encodes the binding domain and simultaneously to deduce the amino acid sequence of the binding domain. Nucleotide sequence analysis can be carried out by any method



known in the art, including but not limited to the method of Maxam and Gilbert (1980, Meth. Enzymol. 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No. 4,795,699; Sequenase™, U.S. Biochemical Corp.), or Taq polymerase, or use of an automated DNA sequenator (e.g., Applied Biosystems, Foster City, CA).

Alternatively, syngenes encoding binding domains are identified by a method comprising identifying their encoded binding domain protein (and/or peptide) which binds to a ligand of choice, comprising: (a) generating a library of vectors expressing a plurality of heterofunctional fusion proteins comprising (i) a binding domain encoded by a double stranded oligonucleotide comprising unpredictable nucleotides in which the unpredictable nucleotides are arranged in one or more contiguous sequences, wherein the total number of unpredictable nucleotides is greater than or equal to about 60 and less than or equal to about 600 and the contiguous sequences are flanked by invariant residues designed to encode amino acids that confer a desired structure to the binding domain of the expressed heterofunctional fusion protein, and, optionally, (ii) an effector domain encoded by an oligonucleotide sequence encoding a protein or peptide that enhances expression or detection of the binding domain; and (b) screening the library of vectors by contacting the plurality of heterofunctional fusion proteins with the ligand of choice under conditions conducive to ligand binding and isolating the heterofunctional fusion protein which binds to the ligand. Additionally, the methods of the invention further comprise determining the nucleotide sequence encoding the binding domain of the heterofunctional fusion protein identified to

deduce the amino acid sequence of the binding domain. Most important, however, the present invention determines the nucleotide sequence encoding the binding domain of the heterofunctional fusion protein  
5 identified to identify the syngene to encode said binding domain.

Once a suitable random peptide library has been constructed (or otherwise obtained), the library is screened to identify peptides having binding  
10 affinity for a ligand of choice. Screening the libraries can be accomplished by any of a variety of methods known to those of skill in the art. See, e.g., the following references, which disclose screening of peptide libraries: Parmley and Smith,  
15 1989, Adv. Exp. Med. Biol. 251:215-218; Scott and Smith, 1990, Science 249:386-390; Fowlkes et al., 1992; BioTechniques 13:422-427; Oldenburg et al., 1992, Proc. Natl. Acad. Sci. USA 89:5393-5397; Yu et al., 1994, Cell 76:933-945; Staudt et al., 1988,  
20 Science 241:577-580; Bock et al., 1992, Nature 355:564-566; Tuerk et al., 1992, Proc. Natl. Acad. Sci. USA 89:6988-6992; Ellington et al., 1992, Nature 355:850-852; U.S. Patent No. 5,096,815, U.S. Patent No. 5,223,409, and U.S. Patent No. 5,198,346, all to  
25 Ladner et al.; and Rebar and Pabo, 1993, Science 263:671-673.

If the libraries are expressed as fusion proteins with a cell surface molecule, then screening is advantageously achieved by contacting the vectors  
30 with an immobilized target ligand and harvesting those vectors that bind to said ligand. Such useful screening methods, designated "panning" techniques are described in Fowlkes et al., 1992, BioTechniques 13:422-427. In panning methods useful to screen the  
35 libraries, the target ligand can be immobilized on plates, beads, such as magnetic beads, sepharose, etc., or on beads used in columns. In particular

embodiments, the immobilized target ligand can be "tagged", e.g., using such as biotin, 2-fluorochrome, e.g. for FACS sorting.

5 In one embodiment, presented by way of example but not limitation, screening a library of phage expressing random peptides on phage and phagemid vectors can be achieved as follows using magnetic beads. Target ligands are conjugated to magnetic beads, according to the instructions of the  
10 manufacturers. To block non-specific binding to the beads, and any unreacted groups, the beads are incubated with excess bovine serum albumin (BSA). The beads are then washed with numerous cycles of suspension in phosphate buffered saline (PBS; 137 mM  
15 NaCl, 2.7 mM KCl, 4.3 mM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1.4 mM  $\text{KH}_2\text{PO}_4$ , pH 7.3) with 0.05% Tween ® 20 and recovered with a strong magnet along the sides of a plastic tube. The beads are then stored with refrigeration, until needed.

In the screening experiments, an aliquot of  
20 a library is mixed with a sample of resuspended beads. The tube contents are tumbled at 4°C for 1-2 hrs. The magnetic beads are then recovered with a strong magnet and the liquid is removed by aspiration. The beads are then washed by adding PBS-0.05% Tween ® 20,  
25 inverting the tube several times to resuspend the beads, and then drawing the beads to the tube wall with the magnet. The contents are then removed and washing is repeated 5-10 additional times. 50 mM glycine-HCl (pH 2.0), 100 µg/ml BSA solution are added  
30 to the washed beads to denature proteins and release bound phage. After a short incubation time, the beads are pulled to the side of the tubes with a strong magnet and the liquid contents are then transferred to clean tubes. 1 M Tris-HCl (pH 7.5) or 1 M  $\text{NaH}_2\text{PO}_4$  (pH  
35 7) is added to the tubes to neutralize the pH of the phage sample. The phage are then diluted, e.g.,  $10^{-3}$  to  $10^{-6}$ , and aliquots plated with *E. coli* DH5αF' cells

to determine the number of plaque forming units of the sample. In certain cases, the platings are done in the presence of XGal and IPTG for color discrimination of plaques (i.e., lacZ+ plaques are blue, lacZ- 5 plaques are white). The titer of the input samples is also determined for comparison (dilutions are generally  $10^{-6}$  to  $10^{-9}$ ).

Alternatively, as yet another non-limiting example, screening a library of phage expressing 10 random peptides can be achieved by panning using microtiter plates. Target ligand is diluted, e.g., in 100 mM NaHCO<sub>3</sub>, pH 8.5 and a small aliquot of ligand solution is adsorbed onto wells of microtiter plates (e.g. by incubation overnight at 4°C). An aliquot of 15 BSA solution (1 mg/ml, in 100 mM NaHCO<sub>3</sub>, pH 8.5) is added and the plate incubated at room temperature for 1 hr. The contents of the microtiter plate are flicked out and the wells washed carefully with PBS-0.05% Tween ® 20. The plates are washed free of 20 unbound targets repeatedly. A small aliquot of phage solution is introduced into each well and the wells are incubated at room temperature for 1-2 hrs. The contents of microtiter plates are flicked out and washed repeatedly. The plates are incubated with wash 25 solution in each well for 20 minutes at room temperature to allow bound phage with rapid dissociation constants to be released. The wells are then washed five more times to remove all unbound phage.

30 In a preferred method for recovering the phage bound to the wells, a pH change is used. An aliquot of 50 mM glycine-HCl (pH 2.0), 100 µg/ml BSA solution is added to the washed wells to denature proteins and release bound phage. After 10 minutes at 35 65°C, the contents are then transferred into clean tubes, and a small aliquot of 1 M Tris-HCl (pH 7.5) or 1M NaH<sub>2</sub>PO<sub>4</sub> (pH 7) is added to neutralize the pH of the

phage sample. The phage are then diluted, e.g.,  $10^{-3}$  to  $10^{-6}$  and aliquots plated with *E. coli* DH5 $\alpha$ F' cells to determine the number of the plaque forming units of the sample. In certain cases, the platings are done in the presence of XGal and IPTG for color discrimination of plaques (i.e., lacZ+ plaques are blue, lacZ- plaques are white). The titer of the input samples is also determined for comparison (dilutions are generally  $10^{-6}$  to  $10^{-9}$ ). Alternatively, to recover bound phage, a large volume, approximately 100  $\mu$ l, of LB+ ampicillin is added to each well and the plate is incubated at 37°C for 2 hr. The bound cells undergo cell division in the rich culture medium and the daughter cells detach from the immobilized targets. The contents of the wells are then transferred to a culture flask that contains ~10 ml LB + ampicillin. When the cells are at log-phase, inducer is added again to the culture to generate more of the encoded proteins. These cells are then harvested by centrifugation and rescreened.

By way of another example, the libraries expressing random peptides as a surface protein of either a vector or a host cell, e.g., phage or bacterial cell, can be screened by passing a solution of the library over a column of a ligand immobilized to a solid matrix, such as sepharose, silica, etc., and recovering those phage that bind to the column after extensive washing and elution.

By way of yet another example, weak binding library members can be isolated based on retarded chromatographic properties. According to one mode of this embodiment for screening, fractions are collected as they come off the column, saving the trailing fractions (i.e., those members that are retarded in mobility relative to the peak fraction are saved). These members are then concentrated and passed over the column a second time, again saving the retarded

fractions. Through successive rounds of chromatography, it is possible to isolate those that have some affinity, albeit weak, to the immobilized ligand. These library members are retarded in their mobility because of the millions of possible ligand interactions as the member passes down the column. In addition, this methodology selects those members that have modest affinity to the target, and which also have a rapid dissociation time. If desired, the oligonucleotides encoding the binding domain selected in this manner can be mutagenized, expressed and rechromatographed (or screened by another method) to discover improved binding activity.

According to another example, homobifunctional (e.g., DSP, DST, BSOCOES, EGS, DMS) or heterobifunctional (e.g., SPDP) cross-linking agents can be used in combination with any of the above methods, to promote capture of weak binding members; these cross-linkers should be reversible, with a treatment (i.e., exposure to thiols, base, periodate, hydroxylamine) gentle enough not to disrupt members structure or infectivity, to allow recovery of the library member. The elution reagents can be removed by dialysis (i.e., dialysis bag, Centricon/Amicon microconcentrators).

According to another alternative method, screening a library of can be achieved using a method comprising a first "enrichment" step and a second filter lift step as follows.

Random peptides from an expressed library capable of binding to a given ligand ("positives") are initially enriched by one or two cycles of panning or affinity chromatography, as described above. The goal is to enrich the positives to a frequency of about  $> 1/10^5$ . Following enrichment, a filter lift assay is conducted. For example, approximately  $1-2 \times 10^5$  phage, enriched for binders, are added to 500  $\mu$ l of log phase

*E. coli* and plated on a large LB-agarose plate with 0.7% agarose in broth. The agarose is allowed to solidify, and a nitrocellulose filter (e.g., 0.45  $\mu$ ) is placed on the agarose surface. A series of registration marks is made with a sterile needle to allow re-alignment of the filter and plate following development as described below. Phage plaques are allowed to develop by overnight incubation at 37 °C (the presence of the filter does not inhibit this process). The filter is then removed from the plate with phage from each individual plaque adhered *in situ*. The filter is then exposed to a solution of BSA or other blocking agent for 1-2 hours to prevent non-specific binding of the ligand (or "probe").

The probe itself is labeled, for example, either by biotinylation (using commercial NHS-biotin) or direct enzyme labeling, e.g., with horse radish peroxidase (HRP) or alkaline phosphatase. Probes labeled in this manner are indefinitely stable and can be re-used several times. The blocked filter is exposed to a solution of probe for several hours to allow the probe to bind *in situ* to any phage on the filter displaying a peptide with significant affinity to the probe. The filter is then washed to remove unbound probe, and then developed by exposure to enzyme substrate solution (in the case of directly labeled probe) or further exposed to a solution of enzyme-labeled avidin (in the case of biotinylated probe). In a preferred method, an HRP-labeled probe is detected by ECL western blotting methods (Amersham, Arlington Heights, IL), which involves using luminol in the presence of phenol to yield enhanced chemiluminescence detectable by brief exposure of film by autoradiography, in which the exposed areas of film correspond to positive plaques on the original plate. Where an enzyme substrate is used, positive phage plaques are identified by localized deposition of

colored enzymatic cleavage product on the filter which corresponds to plaques on the original plate. The developed filter or film, as the case may be, is simply realigned with the plate using the registration marks, and the "positive" plaques are cored from the agarose to recover the phage. Because of the high density of plaques on the original plate, it is usually impossible to isolate a single plaque from the plate on the first pass. Accordingly, phage recovered from the initial core are re-plated at low density and the process is repeated to allow isolation of individual plaques and hence single clones of phage.

Successful screening experiments are optimally conducted using 3 rounds of serial screening. The recovered cells are then plated at a low density to yield isolated colonies for individual analysis. The individual colonies are selected and used to inoculate LB culture medium containing ampicillin. After overnight culture at 37°C, the cultures are then spun down by centrifugation. Individual cell aliquots are then retested for binding to the target ligand attached to a solid support. Binding to other supports, having attached thereto a non-relevant ligand, can be used as a negative control.

One important aspect of screening the libraries is that of elution. For clarity of explanation, the following is discussed in terms of TSAR expression by phage; however, it is readily understood that such discussion is applicable to any system where the random peptide is expressed on a surface fusion molecule. It is conceivable that the conditions that disrupt the peptide-target interactions during recovery of the phage are specific for every given peptide sequence from a plurality of proteins expressed on phage. For example, certain interactions may be disrupted by acid pH's but not by



basic pH's, and vice versa. Thus, it may be desirable to test a variety of elution conditions (including but not limited to pH 2-3, pH 12-13, excess target in competition, detergents, mild protein denaturants, urea, varying temperature, light, presence or absence of metal ions, chelators, etc.) and compare the primary structures of the TSAR proteins expressed on the phage recovered for each set of conditions to determine the appropriate elution conditions for each ligand/TSAR combination. Some of these elution conditions may be incompatible with phage infection because they are bactericidal and will need to be removed by dialysis (i.e., dialysis bag, Centricon/Amicon microconcentrators).

The ability of different expressed proteins to be eluted under different conditions may not only be due to the denaturation of the specific peptide region involved in binding to the target but also may be due to conformational changes in the flanking regions. These flanking sequences may also be denatured in combination with the actual binding sequence; these flanking regions may also change their secondary or tertiary structure in response to exposure to the elution conditions (i.e., pH 2-3, pH 12-13, excess target in competition, detergents, mild protein denaturants, urea, heat, cold, light, metal ions, chelators, etc.) which in turn leads to the conformational deformation of the peptide responsible for binding to the target.

According to another alternative method in which the TSARs contain a linker region between the binding domain and the effector domain, particular TSAR libraries can be prepared and screened by: (1) engineering a vector, preferably a phage vector, so that a DNA sequence encodes a segment of Factor Xa (or Factor Xa protease cleavable peptide) and is present adjacent to the gene encoding the effector domain,

e.g., the pIII coat protein gene; (2) construct and assemble the double stranded synthetic oligonucleotides as described above and insert into the engineered vector; (3) express the plurality of vectors in a suitable host to form a library of vectors; (4) screen for binding to an immobilized ligand; (5) wash away excess phage; and (6) treat the entire library with Factor Xa protease. The particle will be uncoupled from the peptide-ligand complex and can then be used to infect bacteria to regenerate the particle with its full-length pIII molecule for additional rounds of screening. This alternative embodiment advantageously allows the use of universally effective elution conditions and thus allows identification of phage expressing TSARs that otherwise might not be recovered using other known methods for elution. To illustrate, using this embodiment, exceptionally tight binding TSARs could be recovered.

In a particular embodiment of the present invention, TSAR libraries are screened with oligonucleotides to identify syngenes encoding binding domains specific for particular DNA sequences. When the library is screened against oligonucleotide targets, binding is done with high levels of salts, e.g., to maximize the hydrophobic interactions that are characteristic of specific protein/DNA interactions and to minimize non-specific interactions. Non-specific protein/DNA interactions are generally electrostatic and can be reduced by high salt concentrations that cause saturation of the charges on the protein and DNA.

#### 5.2.1. SYNGENES THAT REGULATE TRANSCRIPTION

Syngene-encoded sequences that can be used to regulate transcription are identified by screening the desired random peptide library for binding to a

ligand of choice, where the ligand of choice is (or comprises) a nucleic acid sequence (transcriptional regulatory site) that regulates transcription; a transcription factor that binds to a transcriptional regulatory site and thereby regulates transcription; or a protein binding partner of the transcription factor, which binds to the transcription factor and thereby inhibits the transcription factor's binding to the transcriptional regulatory site.

It is well known that all eukaryotic genes are regulated by proteins known as transcription factors. In general, for each gene, several factors are utilized to appropriately regulate the gene. A large subset of these factors bind directly to DNA sequences near the transcriptional start site of the gene. Most of the sites that bind a given transcription factor have a great deal of similarity among themselves. The nucleotide sequences found in common among all the sites for a given factor are called that factor's "consensus sequence." For example, the consensus sequence for NF-IL6 (also known as C/EBP) is 5'T(T/G)NNGNAA(T/G)3' (SEQ ID NO: 59) (Zhang et al., 1994, Proc. Natl. Acad. Sci. USA 91:2225-2229). The consensus sequence for NF- $\kappa$ B is 5'GGGRNNYYCC (SEQ ID NO: 60) (Okamoto et al., 1993, J. Biol. Chem. 269:8582-8589). The consensus sequence for GATA-1, GATA-2, and GATA-3 is 5'(T/A)GATA(G/A) (Zon et al., 1991, Proc. Natl. Acad. Sci. USA 88:10642-10646).

In addition, certain transcription factors are not single proteins, but are multimers composed of the products of different genes. For instance, the AP-1 transcription site is activated by heterodimers of c-jun and c-fos. Moreover, there are several members of the c-jun family. For the NF- $\kappa$ B family, the transcriptional activator may be composed of either homo- or heterodimers of members of the NF- $\kappa$ B

family, including p65 (also known as Rel A) and p50. Different NF- $\kappa$ B sites may be activated preferentially by different combinations of p65 and/or p50 (Muroi et al., 1993, J. Biol. Chem. 268:19534-19539; Zabel et al., 1991, J. Biol. Chem. 266:252-260; Baldwin and Sharp, 1988, Proc. Natl. Acad. Sci. USA 85:723-727; Nakayama, et al., 1992, Mol. Cell. Biol 12:1736-1746; Fujita et al., 1992, Genes and Development 6:775-787). Thus, different genes may be regulated differentially by specific combinations of transcription factor homologs.

The activity of some transcription factors is regulated by proteins that bind to and thus inhibit the factor's activity. For example, the activity of NF- $\kappa$ B can be inhibited by binding to the protein I $\kappa$ B (Kerr et al., 1992, Current Opin. Cell. Biol. 4:496-501). The activity of the AP-1 transcription factor can be inhibited by binding of the factor IP-1 (Auwerx and Sassone-Corsi, 1991, Cell 64:983-993). Often, as in the case of I $\kappa$ B, the inhibitory protein is located in the cytoplasm, where its binding to the transcription factor sequesters that factor away from the nucleus, thus preventing the factor from activating transcription. Other transcription factors that can be found in the cytoplasm, presumably in inactive form, include: the glucocorticoid receptor, the yeast transcription factor SW15, and NF-AT (a transcription factor for certain lymphokines and cytokines).

Alternative splicing may give rise to several versions of a transcription factor. Some of these versions may inhibit rather than activate transcription, and the interaction of different versions of the same factor may produce a complex pattern of regulation (Foulkes and Sassone-Corsi, 1992, Cell 68:411-414).

A given gene may be regulated by one or more different transcription factors in a tissue or cell type specific manner. For instance, it has been demonstrated that at least three cis-elements (AP-1, C/EBP (also known as NF-IL6) and NF- $\kappa$ B-like sites) are involved in IL-8 gene activation (Okamoto et al., 1994, J. Biol. Chem. 269:8582-8589; Mukaida et al., 1990, J. Biol. Chem. 265:21128-21133; Mahe et al., 1991, J. Biol. Chem. 266:13759-13763; Yasumoto et al., 1992, J. Biol. Chem. 267:22506-22511). It was concluded that the relative importance of the 3 elements varied in a cell type-specific manner. In the human fibrosarcoma cell line 8387, the C/EBP and NF- $\kappa$ B sites are essential, but the AP-1 site is not (Mukaida et al., 1990, J. Biol. Chem. 265:21128-21133; Mahe et al., 1991, J. Biol. Chem. 266:13759-13763). In contrast, in MKN-45 cells, derived from human gastric cancer, the AP-1 and NF- $\kappa$ B sites are important, but the C/EBP site is unnecessary (Yasumoto et al., 1992, J. Biol. Chem. 267:22506-22511).

Transcription factors have been shown to be modular proteins. That is, they are composed of more or less discrete domains that perform more or less discrete functions. For example, most transcription factors possess a DNA binding domain that mediates binding to a transcriptional regulatory site. Also common are transcriptional activator signals (TASs). TASs are domains that, when bound to a transcriptional regulatory site of a gene (by virtue of being linked to a DNA binding domain), mediate the transcriptional activation of the gene. Another common domain found in transcription factors is a domain that permits the dimerization of the factor. The dimerization is sometimes homodimerization, sometimes heterodimerization. An example of such a dimerization domain would be a leucine zipper or a helix-loop-helix domain.

NF- $\kappa$ B is a transcription factor that is a heterodimer consisting of p50 and p65 subunits (Lenardo and Baltimore, 1989, Cell 58:227-229; Baeuerle, 1991, Biochem. Biophys. Acta 1072:63-80).

- 5 IF- $\kappa$ B (or I $\kappa$ B) is a cytoplasmic protein that binds to and specifically inhibits NF- $\kappa$ B (Baeuerle and Baltimore, 1988, Science 242:540-546; Baeuerle & Baltimore, 1989, Genes & Dev. 3:1689-1698). Free NF- $\kappa$ B migrates to the nucleus where it binds to the  $\kappa$ B
- 10 site in the enhancer/promoter of various genes, such as the T cell receptor  $\beta$  chain, interleukin-2 receptor  $\alpha$  chain, myosin heavy chain class I, and cytokine genes such as those for beta-interferon, GM-colony stimulating factor (CSF), G-CSF, interleukin-2, and
- 15 tumor necrosis factor- $\alpha$  and - $\beta$ , thereby regulating their expression (Lenardo and Baltimore, 1989, Cell 58:227-229; Baeuerle, 1991, Biochem. Biophys. Acta 1072:63-80).

- Binding sites for the transcription factors
- 20 NF- $\kappa$ B and NF-IL6 have been shown to be important in the regulation of a wide variety of genes that are involved in health and disease. For example, it has been suggested that derangements in the control of expression of the IL-8 gene may be involved in the
- 25 pathogenesis of several inflammatory diseases (Okamoto et al., 1994, J. Biol. Chem. 269:8582-8589). An NF- $\kappa$ B binding site in the promoter of the IL-8 gene has been shown to be involved in the transcriptional control of the IL-8 gene (*id.*). The same NF- $\kappa$ B site has been
- 30 shown to be one of the targets through which the immunosuppressant FK506 acts (*id.*). The ability to regulate this site through the use of syngenes encoding highly specific DNA binding domains that bind to the site would be of great use in the management of
- 35 those inflammatory disease in which IL-8 is involved.

The GATA family of transcription factors are a group of zinc finger DNA binding proteins that

recognize related binding sites that contain the core sequence GATA. The known members of this family are GATA-1, GATA-2, GATA-3, GATA-GT1, and GATA-GT2 (Zon et al., 1991, Proc. Natl. Acad. Sci. USA 88:10642-10646; 5 Maeda, 1994, J. Biochem. 115:6-14). GATA-1, GATA-2, and GATA-3 are expressed primarily in hematopoietic tissues such as the erythroid, megakaryocyte, T cell, and mast cell lineages. Gene knockout studies have shown that GATA-1 is required for erythroid 10 development in mice (Pevny et al., 1991, Nature 349:257-260). GATA-2 is found in non-hematopoietic cells such as brain, liver, and endothelial cells (Dorfman et al., 1992, J. Biol. Chem. 267:1279-1285).

GATA-GT1 and GATA-GT2 are found in gastric 15 parietal cells. They bind to a sequence motif in the 5' upstream regions of both the H<sup>+</sup>/K<sup>+</sup>-ATPase  $\alpha$  and  $\beta$  subunit genes. GATA-GT1 and GATA-GT2 may be involved in gastric specific transcriptional regulation of many proteins (Maeda, 1994, J. Biochem. 115:6-14).

20 The Ets family of transcription factors includes at least 30 different DNA binding proteins in species as evolutionarily distant as *Drosophila* and humans. These proteins show pronounced amino acid sequence similarities in an approximately 84 amino 25 acid region that corresponds to their DNA binding domains. They recognize distinct but related DNA binding sites that are about ten nucleotides long and that share a common, short motif - (C/A)GGA(A/T). The specificity of binding of the individual Ets proteins 30 is determined mainly by the other nucleotides in the binding site.

Numerous genes have been shown to possess functional Ets binding sites in their promoters (Waslyk, 1993, Eur. J. Biochem. 211:7-18). One 35 particularly interesting example of such a gene is the T cell antigen receptor  $\alpha$  chain gene (TCR $\alpha$ ). It has been shown that mutation of this Ets binding site

abolishes the transcription of a transfected TCR $\alpha$  gene in the Jurkat T cell line. It has also been shown that this site binds the product of the Ets-1 proto-oncogene (Ho et al., 1990, Science 250:814-818). Ets-1 expression is developmentally regulated in parallel with the expression of the TCR $\alpha$  gene, leading to speculation that Ets-1 is the controlling factor in TCR $\alpha$  expression (Leiden, 1993, Ann. Rev. Immunol. 11:539-570).

10

5.2.1.1. SYNGENES ENCODING PROTEINS  
THAT BIND TO DNA SEQUENCES THAT  
REGULATE TRANSCRIPTION

In one embodiment, the invention involves the identification of syngene-encoded peptide sequences that bind to nucleic acid sequences that regulate (promote or alternatively inhibit) transcription (such nucleic acid sequences being collectively referred to as transcriptional regulatory sites). Such transcriptional regulatory sites are preferably those situated *in vivo* in or near the sequence of a preselected gene, the product of which is involved in health or disease. In a specific embodiment, the transcriptional regulatory site is a sequence recognized by a transcription factor. Oligonucleotides containing those transcriptional regulatory sites are synthesized and are used to screen a random peptide library. In this way, peptides are found that specifically bind the transcriptional regulatory sites. In addition, by screening against the appropriate transcriptional regulatory sites, a synthetic binding domain can be identified that specifically regulates the transcription of one or more genes of choice, but not of all the genes whose transcription is regulated by a natural transcription factor which recognizes the same transcriptional regulatory site as does the syngene-encoded binding

35



domain. Such a binding domain, that specifically regulates one or more but not all of a family of transcriptional regulatory sites (a family referring to those transcriptional regulatory sites regulated by an identical natural transcription factor), is referred to herein as a highly specific DNA binding domain (HSDB). HSDBs encoded by syngenes are useful in enhancing or repressing the transcriptional activity of only one or a selected number of a preselected class of genes.

Since the specific nucleotide sequence for a particular example of a given transcriptional regulatory site is often unique, or nearly so, the invention provides for the isolation of HSDBs that bind specifically to a particular transcriptional regulatory site within or near a particular gene. In a specific embodiment, a syngene product is targeted to the nucleus of a cell via a nuclear localization signal (NLS) where it binds to a specific transcriptional regulatory site within a gene whose activity it is desired to regulate for therapeutic reasons. Such binding interferes with the binding of a natural transcription factor to that site and thus blocks the transcription of the preselected gene. The syngene product in this case would not contain a transcription activating signal (TAS).

In a similar fashion, it is possible to incorporate a TAS into the syngene product plus NLS. This will target the TAS to a transcriptional regulatory site of a particular preselected gene within a cell and thus activate that gene selectively. The site at which the syngene product binds could be a site that is naturally involved in the activation of the transcription of the preselected gene. Alternatively, the site could be a site that is not naturally involved in the transcriptional regulation of the gene.

The invention encompasses a method for functionally identifying protein structural domains that can be combined with other functional domains to effect the production of a unique, non-naturally occurring protein that can act at specific cis-acting negative or positive transcriptional regulatory sites within eukaryotic cells. The invention is further directed to the syngenes encoding such a protein as well as to the uses of such a syngene. This invention in a particularly preferred embodiment is directed to the repression and activation of genes regulated, at least in part, by the transcriptional factors NF- $\kappa$ B and NF-IL6 (also known as C/EBP). In particular, this embodiment is directed to the genes for IL-6 and IL-8, whose products are cytokines that respond to different inflammatory signals and are regulated by both NF- $\kappa$ B and NF-IL6. In addition, regulation of the gene for the HLA class I locus, which responds to NF- $\kappa$ B, but not NF-IL6, is a subject of this invention.

In a preferred embodiment, the transcriptional regulatory site in the vicinity of a preselected gene to which the encoded syngene product binds (either for the purpose of activating or repressing transcription of the preselected gene) is an NF- $\kappa$ B site (see Table 6, below). In another preferred embodiment, the site to which the encoded syngene peptide binds is an NF-IL6 site.

In another method of regulating transcription, the encoded syngene product binds to a protein that is a transcriptional regulator (i.e. that binds to DNA and regulates transcription). For example, the encoded syngene product can bind to the NF- $\kappa$ B protein and thereby prevent the NF- $\kappa$ B protein from activating transcription.

In another method of activating transcription, the encoded syngene product need not be localized to the nucleus. In this method, the product

of the syngene functions in the cytoplasm where it binds to an endogenous binding partner of a transcriptional regulator. In one embodiment, the syngene product activates transcription of a  
5 preselected gene by binding to a transcription factor inhibitor that is involved in the inhibition of transcription of the preselected gene. By way of example, such a transcription factor inhibitor is IF- $\kappa$ B.

10 In a preferred aspect of the invention, the syngenes encode a binding domain that is highly specific for the DNA binding site of a transcription factor. Such syngenes encode proteins containing  
15 HSDBs that are identified by screening a random peptide library for binding to a target ligand, in which the ligand is a nucleic acid that contains a binding site recognized by the transcription factor. HSDBs will show greater specificity for a given  
20 transcription factor binding site than does the natural transcription factor that binds to that site. An HSDB selectively binds a single or limited examples of a given transcription factor's binding site from a single gene or a limited number of genes, whereas the natural transcription factor will bind to sites in  
25 additional genes.

Transcription factor binding sites are known for a very large number of genes and are suitable for use as target ligands. For example, the following  
30 tables represent examples of known transcription factors and their respective binding sites that can be used in the methods of the present invention to isolate binding domains specific for those factors or sites. Sequences encoding such binding domains could  
35 be built into syngene constructs that would be useful for the modulation of the activity of genes whose transcription is dependent upon the binding of transcription factors to those sites.

Table 1\*

## Biochemical Characteristics of DNA-Binding Proteins That Regulate Transcription in Early B Cells

Protein	Alternative names	Length (amino acids)	DNA-binding domain	Dimerization <sup>1</sup>	Binding site <sup>2</sup>	Potential targets
BSAP	Pax-5, EBB-1, SαBP, NF-HB complex 3 NFSμ-B1	391	Paired	Monomer	A <sub>G</sub> ---CA-TG---GCG <sub>GA</sub> ACCA <sub>CG</sub>	CD19, VpreB, λ5, b1k, Iε promoters; 5'Sy2a; 5'Sα. 3'Ca enhancer
EBF	B-Lyf, OLF-1	591	EBF	Homodimers require HLH helix 2 like repeats	A <sub>T</sub> TCCCNNGGGA <sub>A</sub> T <sub>T</sub>	mb-1 promoter, olfactory neuron-specific genes
PU.1	Spi-1	264	ETS	Monomer	TG <sub>AA</sub> GGGAAGT	Igμ, Igκ3', Igλ enhancers, mb-1 and J chain promoters
Ets-1	c-Ets-1	440	ETS	Monomer	A <sub>C</sub> CGGA <sub>T</sub> AGT <sub>G</sub> A	Igμ, Igκ3' enhancers, mb-1, TdT, λ5 promoters
Fli-1		451	ETS	Monomer	GG <sub>AA</sub> CAGGAAGTG	Igμ enhancer, TdT, λ5, c-myc promoters
Erg-3		?	ETS		GCAGGAGCA	Igμ enhancer

BCF	BCF-1, BCF-2	?	bHLH	Homo- and hetero- dimers of <i>E2A</i> and <i>E2-2</i> gene products	G <sup>C</sup> A <sup>A</sup> NTGTG <sup>T</sup> G	Ig $\mu$ enhancer, MHC class II enhancer
N-myc		456	bHLH-zip	Heterodimers with <i>E2A</i> gene products or Max	CA <sub>1</sub> GTG	c-myc promoter?
IRF-2	ISGF-1	349	IRF	nd	GAAA <sup>GT</sup> CC <sup>GT</sup> GAAACC	Interferon-responsive promoter, e.g. <i>II-7R</i>
Ikaros	LyF-1	431	5CX <sub>1</sub> CX <sub>12</sub> HX <sub>3</sub> H zinc fingers	Monomer	TT <sup>T</sup> TGGG <sup>G</sup> GA <sup>A</sup> A	<i>Tdt</i> , $\lambda$ 5, <i>VpreB</i> , <i>Ick</i> promoters

1 Dimerization status in solution.

2 The SEQ ID NO. for the binding site of BSAP is SEQ ID NO: 61, for that of EBF is SEQ ID NO: 62, for that of Fli-1 is SEQ ID NO: 63, and for that of IRF-2 is SEQ ID NO: 64.

\* From Hagman and Grosschedl, 1994, Current Opin. in Immunol. 6:222-230

For example, Table 1 shows that a DNA binding site for the Ets-like transcription factor Fli-1 is present in the promoter of the c-myc proto-oncogene. C-myc has been shown to be involved in the development of a wide variety of cancers (Marcu et al., 1992, Ann. Rev. Biochem. 61:809-860; Cole, 1986, Ann. Rev. Genet. 20:361-384). The ability to inhibit the activity of c-myc would be expected to have utility in the treatment of those cancers in which c-myc is upregulated.

Table 2\*  
DNA Sequences Recognized by Eukaryotic  
DNA-Binding Transcriptional Factors

Transcription Factor	Sequence Bound	SEQ ID NO.	Tissue Specificity
API	TGA <sup>(C/C)</sup> T <sup>(C/A)</sup> A		Ubiquitous
ATF	TGACG <sup>(C/T)</sup> <sup>(C/A)</sup> <sup>(C/A)</sup>		Ubiquitous
CRE-BO1	GGGGG		Erythrocytes
Ets-1	<sup>(C/C)</sup> <sup>(A/C)</sup> GGG <sup>(A/T)</sup> G <sup>(T/C)</sup>		B cells Resting T cells
GATA-1	<sup>(T/A)</sup> GATAR		Erythroid cells Megakaryocytes Mast Cells
HIVEN86A	TGGGGATTCCCA	65	Activated T cells B-Lymphocytes
MBF-1	<sup>(C/T)</sup> TAAAAATAA <sup>(C/T)</sup> <sup>(C/T)</sup> <sup>(C/T)</sup>	66	Myocytes
NF-κB	GGGA <sup>(A/C)</sup> TN <sup>(T/C)</sup> CC	67	Ubiquitous
Oct-1	ATGCAAAT		Ubiquitous
Sp1	<sup>(C/T)</sup> <sup>(C/A)</sup> GGC <sup>(C/T)</sup> <sup>(C/A)</sup> <sup>(C/A)</sup> <sup>(C/T)</sup>		Ubiquitous

\* from Gambari and Nastrozzi, 1994, Biochem. Pharm. 47(4):599-610

Table 2 gives the DNA sequences recognized by several ubiquitous transcription factors. Such ubiquitous transcription factor binding sites are ideal subjects for use as targets to select highly specific binding domains for incorporation into syngenes. This is because these factors regulate a wide variety of genes in many cell types. Using the methods of the present invention as disclosed in Section 6.1 and its subsections, one can obtain syngenes that encode products that are able to selectively bind to a particular binding site of a ubiquitous transcription factor in a particular gene. This allows for the regulation of that particular gene without disturbing the regulation of the many other genes that also contain DNA binding sites for the ubiquitous transcription factor.

20

25

30

35

Table 3\*

Transcription Factors Involved in Regulating  
Gene Expression During T-Cell Development and Activation

Function Classification	DNA-Binding Domain	Cognate Sequence**	Candidate Targets	Reference
<b>Lineage determination</b>				
bHLH proteins E1 2/E47 ?	bHLH	CANNTG	TCR $\alpha$ , TCR $\beta$ , TCR $\delta$ , CD3 $\epsilon$	[1;2]
Homeobox proteins LH-2 CBF	Homeobox Homeobox	? (A/T)G(C/T)GG(T/A)(T/A)	? TCR $\alpha$ , TCR $\beta$ , TCR $\gamma$ , TCR $\delta$	[3] [4,5,6,7,9]
Zinc finger proteins GATA-3 Ikaro	C4 zinc finger Hunchback-like zinc finger	(A/T)GATA(G/A) TGGGGGT	TCR $\alpha$ , TCR $\beta$ , TCR $\gamma$ , TCR $\delta$ , CD3 $\delta$	[10-14] [15]
<b>Chromatin structure</b>				
HMG proteins LEF-1 TCF-1 SOX-4 GATA Proteins	HMG box HMG box HMG box See GATA-3 above	(A/T)(A/T)CAAAAG (A/T)(A/T)CAAAAG AACAAG See GATA-3 above	CD3 $\epsilon$ , TCR $\alpha$ , TCR $\beta$ , TCR $\delta$ , CD4 CD3 $\epsilon$ , TCR $\alpha$ , TCR $\beta$ , TCR $\delta$ , CD4 CD3 $\epsilon$ See GATA-3 above	[16,17,18*] [19,20] [21*] [10-14]



Transcriptional activation/repression  
Generalized transcription factors

API	Basic-leucine zipper	TGANTCA	IL-2, GM-CSF	[22,23]
CREB/ATF	Basic-leucine zipper	TGANNTCA	TCR $\alpha$ , TCR $\beta$	[24]
Myb		(T/C)AAC(G/T)G	TCR $\delta$ , CD4	[9,25]
NF- $\kappa$ B/rel		GGGA(A/T)(A/T)(T/T)(C)(C/A)C	IL-2, IL-2R	[26,27]
NFATp		GAGGAAAA	IL-2, GM-CSF, IL-4	[28]
Ets transcription factors				
Ets-1	Ets-domain	(C/A)GGA	TCR $\alpha$ , TCR $\beta$	[8,29]
Elf-1	Ets-domain	(C/A)GGA	GM-CSF, IL-3	[30]
Fli-1	Ets-domain	(C/A)GGA	?	[31]
GABP $\alpha$	Ets-domain	(C/A)GGA	?	
Ets-2	Ets-domain	(C/A)GGA	?	

\* From Leiden and Thompson, 1994, Current Opinion in Immunology 6:231-237

GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin; IL-2R, interleukin-2 receptor; TCR, T-cell receptor.  
" N = any nucleotide. The SEQ ID NO. for the sequence in this column corresponding to NF- $\kappa$ B/rel is SEQ ID NO: 68.

- 1 Emerson, 1993, Curr. Opin. Gen. Dev. 3:265-274
- 2 Olson, 1990, Genes Dev. 4:1454-1461
- 3 Deguchi et al., 1991, N. Biol. 3:353-363
- 4 Speck et al., 1990, 4:233-242
- 5 Wang and Speck, 1992, Mol. Cell. Biol. 12:89-102
- 6 Wang et al., 1993, Mol. Cell. Biol. 13:3324-3339
- 7 Miyoshi et al., 1991, Proc. Natl. Acad. Sci. USA 88:10431-10434
- 8 Wotton et al., 1994, Mol. Cell. Biol. 14:840-850
- 9 Hernandez-Munain and Krangel, Mol. Cell. Biol. 14:473-483
- 10 Tsai et al., 1989, Nature 339:446-451
- 11 Evans and Felsenfeld, 1989, Cell 58:877-885
- 12 Pevny et al., 1991, Nature 349:257-260
- 13 Ho et al., 1991, EMBO J. 10:1187-1192
- 14 Ko et al., 1991, Mol. Cell. Biol. 11:2778-2784

- 15 Georgopoulos et al., 1992, Science 258:808-812  
16 Waterman et al., 1991, Genes Dev. 5:656-669  
17 Travis et al., 1991, Genes Dev. 5:880-894  
18 Carlsson et al., 1993, Genes Dev. 7:2418-2430  
19 Van de Wetering et al., 1991, EMBO J. 10:123-132  
20 Oosterwegel et al., 1991, J. Exp. Med. 173:1133-1142  
21 Van de Wetering et al., 1993, EMBO J. 12:3847-3854  
22 Jain et al., 1992, J. Immunol. 148:1240-1250  
23 Jain et al., 1992, Nature 356:801-804  
24 Leiden, 1993, Annu. Rev. Immunol. 11:539-570  
25 Siu et al., 1992, 12:1591-1604  
26 Blank et al., 1992, Trends Biochem. Sci. 4:135-140  
27 Liou and Baltimore, 1993, Curr. Opin. Cell. Biol. 5:477-487  
28 McCaffrey et al., 1993, Science 262:750-754  
29 Ho et al., 1990, Science 250:814-818  
30 Wang et al., 1992, J. Exp. Med. 175:1391-1399  
31 Ben-David et al., 1991, Cell 61:85-99

Table 3 shows a number of transcription factors, the DNA sequences they bind to, and the genes they regulate during T cell development and activation. These DNA sequences are attractive targets for use in obtaining highly specific DNA binding domains that recognize them. Because of the well known crucial role that T cells play in the immune system, syngenes that encode products that are able to regulate the activity of the genes listed in Table 3 would be expected to have a variety of uses in the treatment of disease.

15

20

25

30

35

Table 4\*

Transcription Factors That Regulate TCR Gene Expression

Transcription Factor	Promoter Enhancer Binding Sites	Molecular Mass	DNA-binding Domain	Recognition Sequence	Lineage Specificity	References
CREB		43 kDa	Basic domain/ leucine zipper		Ubiquitous	1,2
CRE-BP1/ATF-2/ mXBP	$\alpha$ , $\beta$ , $V\beta$	55 kDa	Basic domain/ leucine zipper	TGACGTCA	Ubiquitous	3,4,5
ATF-4		38 kDa	Basic domain/ leucine zipper		Ubiquitous	3,6
TCF-1	$\alpha$ , $\beta$ , $\delta$	30 kDa	HMG Box	C <sub>GT</sub> C <sub>T</sub> GAA	T-cells	7,8
TCF-1 $\alpha$ /LEF-1		53-57 kDa	HMG Box	T TTT	T-cells/ Pre-B-cells	7,9
Ets-1	$\alpha$ , $\beta$ , $\gamma$ , $\delta$	54-60 kDa	Basic domain/ $\alpha$ helix	A <sub>T</sub> GA <sub>GGG</sub> TG <sub>C</sub> G CCAC	T-cells/B-cells	10
GATA-3	$\alpha$ , $\beta$ , $\delta$	47 kDa	2 Cx zinc fingers	A <sub>GATA</sub> A TG	T-cells/kidney/ brain	11-14
CBF	$\beta$ , $\gamma$ , $\delta$	19-40 kDa	??	CT <sub>GTGG</sub> TTT TCCCA A	Thymus/spleen	15

<sup>1</sup> Hoeffler et al., 1988, Science 242:1430-1433

<sup>2</sup> Gonzalez et al., 1989, Nature 337:749-752

<sup>3</sup> Hai et al., 1989, Genes Dev. 3:2083-2090

<sup>4</sup> Ivashkiv et al., 1990, Mol. Cell. Biol. 10:1609-1621

- 5 Mackawa et al., 1989, EMBO J. 8:2023-2028  
6 Karpinski et al., 1992, Proc. Natl. Acad. Sci. USA 89:4820-4824  
7 Oosterwegel et al., 1991, J. Exp. Med. 173:1133-1142  
8 van de Wetering et al., 1991, EMBO J. 10:123-132  
9 Travis et al., 1991, Genes Dev. 5:880-894  
10 Ho et al., 1990, Science 250:814-818  
11 Ho et al., 1991, EMBO J. 10:1187-1192  
12 Marine and Winoto, 1991, Proc. Natl. Acad. Sci. USA 88:7284-7288  
13 Joulin et al., 1991, EMBO J. 10:1809-1816  
14 Yamamoto et al., 1990, Genes Dev. 4:1650-1662  
15 Wang and Speck, 1992, Mol. Cell. Biol. 12:89-102

\* From Leiden, 1993, Annu. Rev. Immunol. 11:539-70

Table 4 shows a number of transcription factors that could be used as targets to obtain syngenes that encode products that would be useful in regulating the T cell receptor.

5

Table 5*			
Specific Binding Sequences for Ets Family Members			
10			SEQ ID NO:
15	ETS1	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		Polyomavirus PEA3	G C A <u>G G A</u> A G T G 69
		Stromelysin	G C A <u>G G A</u> A G C A 70
			C C A <u>G G A</u> A A T G 71
		MSV LTR	A C A <u>G G A</u> T A T C 72
			A G C <u>G G A</u> A G T G 73
		TCR-alpha	A G A <u>G G A</u> T G T G 74
		TCR-beta	A C A <u>G G A</u> T G T G 75
		Ig $\kappa$ 3' enhancer	T C A <u>G G A</u> T G T G 76
		HTLV-I	G G A <u>G G A</u> A A T G 77
			C C G <u>G G A</u> A G C C 78
			C G C <u>G G A</u> A A T G 79
		<u>Selected sites:</u>	
		Fisher et al., 1991	Pu C C <u>G G A</u> A G T Py 80
			G a C
20		Nye et al., 1992	A C C <u>G G A</u> A G C N 81
			g a T a T
		Woods et al., 1992	A C C <u>G G A</u> A G T T 82
25	ETS2	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		Polyomavirus PEA3	G C A <u>G G A</u> A G T G 83
		Stromelysin	G C A <u>G G A</u> A G C A 84
			C C A <u>G G A</u> A A T G 85
		Woods et al. consensus	C C A <u>G G A</u> A G T G 86
	ERG	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		E74	A C C <u>G G A</u> A G T A 87
	GABP	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		HSV 1CP4	A A C <u>G G A</u> A G C G 88
			A G C <u>G G A</u> A A C C 89
30	PEA3	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		Polyomavirus PEA3	G C A <u>G G A</u> A G T G 90
	ELK1/ SAP1	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		Fos	A C A <u>G G A</u> T G T C 91
		E74	A C C <u>G G A</u> A G T A 92
35	E74	<u>Selected site:</u>	1.2.3.4.5.6.7.8.9.10
		Urness & Thummel 1990	Py C A <u>G G A</u> A G T N 93
	ELF1	<u>Promoter sites:</u>	1.2.3.4.5.6.7.8.9.10
		IL-2 NFAT-1	A G A <u>G G A</u> A A A A 94
		IL-2 NFIL-2B	G G A <u>G G A</u> A A A A 95
		HIV-2 CD3R	A C A <u>G G A</u> A C A G 96

<u>Promoter sites:</u>		1.2.3.4.5.6.7.8.9.10	
PUL	MHCII	A G A G G A A C T T	97
	Ig-Kappa 3' enhancer	T G A G G A A C T G	98

- 5 The 10-nucleotide sequence surrounding the GGA core is shown. The sequences were reported by Wasylyk et al., 1990, Nature 346:191-193 (polyoma virus, ETS1+2), Wasylyk et al., 1991, EMBO J. 10:1127-1134 (stromelysin, Ets 1+2), Nye et al., 1992, Genes Dev. 6:975-990 (MSV), Ho et al., 1990, Science 250:814-818 (TCR  $\alpha$ ), Wang et al., 1992, J. Exp. Med. 175:1391-1399 (TCR  $\beta$ , Ig $\kappa$ ), Gitlin et al., 1991, J. Virol. 65:5513-5523 (HTLV1), Fisher et al., 1991, Oncogene 6:2249-2254 (Ets1-selected site), Nye et al., 1992, Genes Dev. 6:975-990 (Ets1-selected site), Woods et al., 1992, Nucleic Acids Res. 4:699-704 (Ets1-selected site, Ets2), Reddy and Rao, 1991, Oncogene 6:2285-2289 (Erg), Thompson et al., 1991, Science 253:762-768 (GABP), Xin et al., 1992, Genes Dev. 6:481-496 (PEA32), Hipkind et al., 1991, Nature 354:531-534 (Elk1, Fos), Rao and Reddy, 1992, Oncogene 7:65-70 (Elk1, E74), Dalton and Treisman, 1992, Cell 68:597-612 (SAP1, Fos), Urness and Thummel, 1990, Cell 63:47-61 (E74), Wang et al., 1992, J. Exp. Med. 175:1391-1399 (Elf1), Klemsz et al., 1990, Cell 61:113-124 (PUL, MHCII), Pongubala et al., 1992, Mol. Cell. Biol. 12:368-378 (PUL, Ig- $\kappa$  3' enhancer).

- 15 From Wasylyk et al., 1993, Eur. J. Biochem 211:7-18

- 20 Table 5 shows a large number of known Ets family transcription factor binding sites. Among the genes listed in Table 5, genes such as fos (involved in growth control) and interleukin-2 (involved in the control of the immune system) are attractive candidates for regulation by syngene products.

- 25 Examples of NF- $\kappa$ B sites which are suitable for use as a ligand of choice are shown in Table 6 (see Muroi et al., 1993, J. Biol. Chem. 268:19534-19539; Lenardo and Baltimore, 1989, Cell 58:227-229).

30

Table 6

NF- $\kappa$ B Transcriptional  
Regulatory Sites in Genes

	<u>DNA SEQUENCE</u>	<u>SEQ ID NO.</u>	<u>GENE</u>
35	GGGAAATTCC	99	beta-interferon
	GGGATTTTCC	100	interleukin-6
	TGGAATTTCC	101	interleukin-8

	GGGACTTCCC	102	MHC class II, I-E $\alpha^d$
	GGGAATACC	103	GM-colony stimulating factor
5	GGGGAATCTC	104	G-colony stimulating factor
	GGGAATCCTT	105	tumor necrosis factor- $\alpha$ 1 (nucleotide position -850)
	GGGGCTTTCC	106	tumor necrosis factor- $\alpha$ 2 (nucleotide position -510)
10	GGGGCTGCCC	107	tumor necrosis factor- $\alpha$ 3 (nucleotide position -610)
	GGGAATATTC	108	Fc $\gamma$ R
	GGGGATTCCC(C)	109	MHC class I (H2-K $b$ )
15	GGGGATTCCCC	109	HLA Class I genes
	GGGACTTTCC	110	HIV NF- $\kappa$ B
	GGGACTTTCC	110	Mouse Ig $\kappa$ enhancer
	GGGACTTTCC	110	SV40 enhancer
20	GGGACTTTCC	110	CMV(4)
	GGGACTTTCC	110	$\beta_2$ -microglobulin
	GGGACTTTCC	110	serum amyloid A
	GGGGATTTC	111	human Ig $\kappa$ enhancer
	GGGGATTTC	111	CMV(3)
25	GGGNACTTTCC	110	CMV(2)
	GGGATTTTCC	112	interleukin-2(IL-2)
	GGGATTTCTT	113	mouse IL-2 receptor $\alpha$
	GGGAATCTCC	114	human IL-2 receptor $\alpha$
	GGGRNNYYCC*	115	consensus motif
30			

* R = A or G; N = A, G, T or C; Y = C or T
--

For those genes of interest for which transcription factor binding sites are not known, as in the case, for example, of a newly discovered gene, well known methods will allow the determination of transcription factor binding sites that are involved



in the regulation of the gene. Such methods include, for example, comparing the DNA sequences in the 5' region of the gene to the sequences of known transcription factor binding sites, gel shift assays to determine if known transcription factors bind to potential transcription factor binding sites in the gene, and mutagenesis of a suspected binding site followed by assays to determine if the expression of the mutagenized gene has been affected by the mutagenesis.

In an alternative embodiment of the invention, the molecule comprising the transcriptional regulatory site that constitutes the ligand of choice used to screen the random peptide library is a site that is not known to be bound by a protein transcription factor, but which is situated close to (e.g., within 50 nucleotides of) the initiation site of the gene, and which has the ability to regulate transcription of such gene upon binding by a syngene-encoded product. Such a syngene-encoded product would be expected to sterically hinder the ordered process of transcription factor and RNA polymerase binding to this region of the gene that is necessary for transcription to occur.

5.2.1.2. SYNGENES FOR REGULATION OF  
TRANSCRIPTION FACTORS AND  
TRANSCRIPTION FACTOR INHIBITORS

The invention provides a method for identifying a syngene that encodes a protein that binds to a ligand of interest comprising screening a random synthetic peptide library to identify a peptide that binds to a ligand of interest, in which the ligand is (1) a DNA region that functions to regulate transcription; (2) a protein that is a transcriptional regulator that functions by binding to DNA; or (3) a protein modulator that binds to and inactivates the function of a protein transcriptional regulator.

In cases such as that of NF- $\kappa$ B and IF- $\kappa$ B (Karr et al., 1992, Current Opinion in Cell Biology 4:496-501) or of AP-1 and IP-1 (Auwerx and Sassone-Corsi, 1991, Cell 64:983-993), the existence of a transcription factor and an inhibitor of that transcription factor affords the opportunity for regulating the transcription of genes that depend upon the transcription factor for their activity. This can be done by developing syngenes that encode binding domains that are specific for either the transcription factor or its inhibitor.

If a random peptide library is screened with a transcription factor as a ligand, binding domains (and their encoding nucleic acids) will be identified that are capable of specifically binding the transcription factor. In that respect, such binding domains mimic transcription factor inhibitors. Such binding domains can function as inhibitors of transcription in much the same way as the natural transcription factor inhibitors do. If such a binding domain were built into a syngene-encoded product lacking a nuclear localization signal, the syngene product encoding such a binding domain would be expected to be found predominantly in the cytoplasm where it would be able to sequester the transcription factor away from the nucleus, thus nullifying the transcription factor's activity. However, it is envisioned that such inhibition can also occur in the nucleus, e.g., if a nuclear localization signal were employed as part of the encoded syngene product.

For example, if the library were screened with the protein NF- $\kappa$ B, binding domains would be identified that could function, when part of a syngene-encoded product, much in the way IF- $\kappa$ B functions. Likewise, when a library is screened with either c-fos or c-jun (the subunits of AP-1), binding

domains would be identified that could be used to create IP-1-like peptides encoded by syngenes.

If random peptide libraries were screened with transcription factor inhibitors according to the invention, binding domains would be identified that, when expressed from a syngene, would bind to the inhibitor, thus preventing it from binding to the transcription factor. This would free the transcription factor, allowing it to migrate into the nucleus and activate transcription. In this way, syngenes encoding binding domains specific for transcription factor inhibitors could be used as activators of transcription.

#### 5.2.2. SYNGENES FOR REGULATION OF APOPTOSIS

Syngenes may be used to control apoptosis. Apoptosis, the process of programmed cell death, is seen in a wide range of organisms where it is involved in the development of the nervous system, the remodeling of limb buds, the destruction of virally infected cells, the clonal deletion of B and T cells in the immune system, and many other processes. Programmed cell death thus appears to be a normal part of development and physiology. For reviews, see Kerr et al., 1972, Br. J. Cancer 26:239-257; Wyllie et al., 1980, Int. Rev. Cytol. 68:251-306; Raff, 1992, Nature 356:397-400; Vaux, 1993, Proc. Natl. Acad. Sci. USA 90:786-789.

Since apoptosis is a normal and important part of many different biological processes, it is important for the health of an organism that apoptosis be properly regulated. It appears that some instances of disease can arise as a result of improper regulation of cell death. For example, Alzheimer's and Parkinson's diseases are associated with the premature death of certain neurons (Jenner, 1989, J. Neurol. Neurosurg. Psychiatry 22-28; Kosik, 1992,

Science 256:780-783); inhibition of programmed cell death may be involved in autoimmune disease (Goldstein, et al., 1991, Immunol. Rev. 121:29-65; Watanabe et al., 1992, Nature 356:314-317; Strasser et al., 1991, Proc. Natl. Acad. Sci. USA 88:8661-8665);  
5 and failure of cell death may be involved in cancer (Korsmeyer, 1992, Blood 80:879-886). Given the importance of apoptosis in health and disease, it would be of great value to have a method of regulating  
10 apoptosis.

There appears to be a family of related proteins involved in the control of apoptosis in mammalian cells. The two best-studied members of this family are bcl-2 and bax. Bcl-2 was first identified  
15 as a gene on human chromosome 18 that was involved in the t(14;18) chromosomal translocations observed in follicular B-cell lymphomas (Tsujimoto et al., 1985, Science 229:1390-1393; Bakhshi et al., 1985, Cell 41:899-906; Cleary and Sklar, 1985, Proc. Natl. Acad. Sci. USA 82:7439-7449).  
20

Together, bcl-2 and bax proteins control the process of apoptosis in response to various stimuli. It has been postulated that the ratio of bcl-2 to bax determines whether a cell will initiate the program of  
25 apoptosis in response to apoptotic stimuli or will ignore such stimuli and continue to grow, or at least survive (Oltvai et al., 1993, Cell 74:609-619). A high ratio of bcl-2 to bax leads to survival; a low ratio to death. The interaction of bcl-2 and bax is  
30 thought to depend on the formation of heterodimers of the proteins encoded by these two genes. A study in which mutations were engineered in two domains of bcl-2 found a correlation between the mutant bcl-2  
proteins' ability to heterodimerize with bax and the  
35 mutant proteins' ability to prevent apoptosis (Yin et al., 1994, Nature 369:321-323).

The two domains in the human bcl-2 protein that appear to interact with bax are amino acids 136-156 and amino acids 187-202 (*ibid.*). By using one of these two domains as a ligand to screen a random peptide library, it should be possible to isolate binding domains that are specific for the bcl-2 protein. These binding domains could be used to make syngenes that would be useful in controlling the process of apoptosis. Alternatively, it is possible to use other regions of the bcl-2 protein, or even the entire protein, to find binding domains that are specific for the bcl-2 protein. Such binding domains could also be used to make syngenes that would be useful in the control of apoptosis. Likewise, one could screen a random peptide library with the bax protein to find binding domains specific for bax. These binding domains would also be useful in making syngenes for use in regulating apoptosis.

As an alternative way of regulating apoptosis, one could also regulate the transcription of either bcl-2 or bax by the use of syngenes directed to a transcription factor binding site that is involved in the regulation of the transcription of those genes. The methods described herein for the use of syngenes to regulate transcription could be easily adapted to regulate the transcription of the bcl-2 gene or the bax gene.

### 5.2.3. OTHER TARGETS FOR SYNGENE PRODUCTS

Although a preferred embodiment of the invention is directed to syngenes whose products bind to transcriptional regulatory sites, other embodiments of the invention encompass a plethora of uses. In particular, it should be emphasized that syngene products are in no way restricted to functioning solely in the nucleus. As discussed below, syngenes

are useful for regulating activities that take place in the cytoplasm, or on the cell's surface, as well.

A syngene product may have one or more of a variety of functional properties. Its product may be a signal transduction inhibitor. Such a signal transduction inhibitor would block the activity of a signal pathway such as that which leads from a cell membrane receptor through various intermediate kinases, phosphatases or other molecules to the nucleus, where transcription of particular genes is affected. An example of such a signal transduction inhibiting syngene would be a syngene the product of which binds to and blocks the function of a tyrosine kinase membrane receptor. Other types of signalling molecules may also be targets of the syngenes of the invention. Examples of such signalling molecules are cytoplasmic kinases (for example, src or raf), adenylyl cyclases, guanylyl cyclases and the like.

Many of the proteins that participate in signalling possess SH2 and/or SH3 domains. These SH2 and SH3 domains are involved in protein/protein interactions that are important for the signalling process (Anderson et al., 1990, Science 250:979-982). It may be that syngenes that encode a protein having binding domains specific for SH2 and/or SH3 domains will be especially useful for regulating the various signalling pathways of cells.

In another embodiment, the syngenes, via expression of their encoded proteins, can be used to block or down-regulate the translation of a specific mRNA. Alternatively, the syngene can be used to stabilize or destabilize a specific RNA. In each of the aforementioned cases, the syngene encodes a protein that incorporates a binding domain that has been selected for binding to a portion of the RNA sequence that is to be down-regulated, stabilized or destabilized.

In another embodiment, a syngene is created whose product mimics the structure of a known hormone receptor but that has a different DNA binding site from that of the known hormone receptor. The DNA  
5 binding site of the known hormone receptor is replaced with a HSDB directed against a DNA binding site that represents a transcriptional regulatory site of a preselected gene. The effect of the use of such a  
10 syngene is that the syngene product (which mimics the hormone receptor) allows the preselected gene to be regulated by the hormone. The hormone binds to the syngene product and the syngene product-hormone complex is translocated to the nucleus of a cell where it binds to the transcriptional regulatory site of the  
15 preselected gene, thus activating or repressing the gene.

Syngenes encoding binding domains directed to cyclins or cyclin dependent kinases can be used to inhibit the progression of cells through the cell  
20 cycle. Syngene products containing binding domains directed to DNA polymerases and associated factors involved in DNA replication can be used to stop or retard DNA synthesis, thus stopping or retarding cell division.

25 Another potential target for syngene products of the present invention is the cellular trafficking machinery that directs or sorts proteins, for example, into the Golgi and to their final destinations. The proper subcellular localization of  
30 many proteins depends on short amino acid sequences present in the proteins. For example, the tetrapeptide KDEL (SEQ ID NO: 116) targets proteins to the lumen of the endoplasmic reticulum. The tetrapeptide YQRL (SEQ ID NO: 117), found in the  
35 membrane protein TGN38 of the trans Golgi network, has been demonstrated to be necessary and sufficient to target membrane proteins to the trans Golgi (Luzio and

Banting, 1993, Trends in Biochem. Sci. 18:395-398). Such peptide targeting sequences are appropriate ligands for use in isolating binding domains for incorporation into syngenes.

5                Syngenes can also be useful in regulating the interactions of cells with the extracellular matrix. Such interactions depend on binding between the components of the matrix and receptors on the surface of cells. One of the most important  
10 components in certain types of matrices is the protein laminin. By interacting with specific receptors on the surface of cells, laminin serves as a bridge between the cells and the matrix (Martin and Timpl, 1987, Ann. Rev. Cell Biol. 5:57-85).

15                The interaction between laminin and its receptor has been demonstrated to be important in the migration of axons and dendrites of neurons. The axons and dendrites, guided by the specific interaction between laminin in the extracellular  
20 matrix and laminin receptors on the cell surfaces of the neurons, migrate along extracellular matrix pathways to their destinations (Bray and Hollenbeck, 1988, Ann. Rev. Cell Biol. 4:43-61).

              By screening a random peptide library for  
25 binding domains that are specific for laminin or its receptor, it is possible to isolate binding domains that can be incorporated into the peptides encoded by syngenes that are useful in modulating the process of axonal or dendrite migration as well as other  
30 processes that depend on the interaction between laminin and its receptor.

              The regulation of the intracellular movement of membranes and their contents in mammalian cells is another area where syngenes may be employed.  
35 At least seven classes of GTPases are thought to be involved in such membrane transport processes as movement of membrane vesicles from the endoplasmic



reticulum (ER) through the Golgi to the plasma membrane (or to intracellular organelles) and endocytosis. The seven classes of GTPases are:

- (1) Sar proteins
- 5 (2) Arf proteins
- (3) Rab proteins
- (4) Rac proteins
- (5) Rho proteins
- (6) heterotrimeric G proteins
- 10 (7) dynamin

All of these proteins have the ability to bind GDP or GTP. They cycle between a state in which they are bound to GDP and a state in which they are bound to GTP. Whether they are bound to GDP or GTP  
15 determines how these proteins will function with the general rule being that they are inactive when bound to GDP and active when bound to GTP.

Sar proteins are believed to be involved in membrane vesicle budding from the ER and transport of  
20 the budded vesicles to the Golgi. A postulated mechanism envisions Sar proteins in their GDP-bound form interacting with a specific receptor on the transitional elements of the ER. Transitional elements are specialized sites for the export of newly  
25 synthesized proteins from the ER. This is followed by the binding of four other proteins (Sec 13, Sec 24, Sec 25, and a 105 kd protein) to the bound Sar to form a nascent vesicle. The nascent vesicle then buds off the ER in a process accompanied by the exchange of the  
30 bound GDP for GTP. This exchange is catalyzed by a Sar specific guanine nucleotide exchange factor (GEF). The budded vesicle then moves to the Golgi where it fuses with the membrane of the *cis* Golgi network in a process that is dependent upon the hydrolysis of the  
35 bound GTP to GDP. This hydrolysis is catalyzed by a Sar specific GTPase activating protein (GAP). See

Figure 2A of Nuoffer and Balch, 1994, Ann. Rev. Biochem. 63:949-990.

Arf proteins are believed to function in a manner similar to Sar proteins. But, whereas Sar proteins are involved in vesicle budding and transport between the ER and the Golgi, Arf proteins are thought to mediate transport between compartments of the Golgi. The mechanism of action of Arf is thought to involve binding of Arf (in its GDP-bound state) to a specific receptor on the membrane of a Golgi vesicle. This is followed by binding to Arf of coat proteins in a process that is dependent upon exchange of the bound GDP for GTP. This exchange is catalyzed by an Arf specific GEF. Movement of the vesicle between Golgi compartments then occurs, followed by docking of the vesicle at its target membrane in the Golgi. Finally, fusion of the vesicle with its target occurs in a process driven by hydrolysis of the bound GTP. This hydrolysis is catalyzed by an Arf specific GAP. See Figure 2B of Nuoffer and Balch, 1994, Ann. Rev. Biochem. 63:949-990.

Sar proteins, Arf proteins, and their associated GEFs and GAPs are suitable targets for the action of syngenes. A random peptide library could be screened for binding domains that are specific for these proteins. Such binding domains could be incorporated into the peptides encoded by syngenes that would be useful for regulating membrane traffic from the ER to the Golgi.

In contrast to Sar and Arf, dynamin does not seem to be involved in transport of membrane vesicles from the ER to the Golgi. Instead, dynamin seems to participate in the receptor mediated endocytotic pathway involving clathrin-coated vesicles. Mutants of dynamin that are unable to bind or hydrolyze GTP block receptor mediated endocytosis in mammalian cells (van der Blik et al., 1993, J. Cell Biol. 122:553-

563; Herskovits et al., 1993, J. Cell Biol. 122: 565-578). It has been proposed that dynamin may be involved in the rapid endocytosis of synaptic vesicles after exocytosis in nerve terminals (Robinson, et al., 1993, Nature 365:163-166). Syngenes containing binding domains specific for dynamin could be used to regulate the process of endocytosis.

Rab proteins are a large class of proteins (more than 30 members have been discovered in mammalian cells alone) that are involved in membrane trafficking. It has been possible to implicate specific members of the Rab class in specific aspects of membrane movement. For example, much evidence indicates that Rab 1 is involved in the early phases of the secretory pathway such as vesicle budding from the ER and vesicle movement from the ER to the *cis* Golgi (Peter et al., 1993, J. Cell Biol. 122:1155-1168; Plutner et al., 1991, J. Cell Biol. 115:31-43; Tisdale et al., 1992, J. Cell Biol. 119:749-761; Plutner, et al., 1990, EMBO J. 10:785-792). Mutated Rab 2 proteins have been shown to be *trans* dominant inhibitors of membrane transport from the ER to the Golgi (Tisdale et al., 1992, J. Cell Biol. 119:749-761). Rab 6 is found in a complex with a 62 kd protein in the membranes of the *trans* Golgi network. Antibodies to Rab 6 (or to the 62 kd protein) inhibited the budding of vesicles from the *trans* Golgi to the plasma membrane, suggesting that Rab 6 may be involved in this process (Jones et al., 1993, J. Cell Biol. 122:775-788). As Section 6.1 and its subsections show, by using the methods of the present invention, one can isolate binding domains that can distinguish between closely related ligands. One can then incorporate those binding domains into syngenes. In this way, it should be possible to make syngenes containing binding domains that are specific for a single member of the Rab family. These syngenes can

then be used to modulate those aspects of membrane trafficking in which that member participates without affecting other aspects.

Evidence supports the involvement of heterotrimeric G proteins in some aspects of membrane trafficking. For example, agonists and antagonists of  $G_o$  (one of the subunits of heterotrimeric G proteins) have been shown to affect export from the ER (Schwaninger et al., 1992, J. Cell Biol. 119:1077-1096). In PC12 and MDCK cells,  $G_{oi}$  inhibits membrane budding from the trans Golgi while  $G_{os}$  promotes budding (Leyte et al., 1992, Trends Cell Biol. 2:91-94; Pimplikar and Simons, 1993, Nature 362:456-458; Barr, et al., 1991, FEBS Lett. 294:239-243). Syngenes containing binding domains specific for the above-mentioned heterotrimeric G proteins would be useful in regulating these aspects of membrane movement.

### 5.3. IDENTIFICATION OF SYNTHETIC SEQUENCES WHICH MEDIATE BINDING

When a peptide from a random peptide library has been identified as a binder for a particular target ligand of interest according to the methods of the invention, it may be useful to determine what region(s) of the expressed peptide sequence is (are) responsible for binding to the target ligand. Such analysis can be conducted at two different levels, i.e., the nucleotide sequence and amino acid sequence levels.

By molecular biological techniques it is possible to verify and further analyze a ligand binding peptide at the level of the oligonucleotides. First, the inserted oligonucleotides can be cleaved using appropriate restriction enzymes and religated into the original expression vector and the expression product of such vector screened for ligand binding to identify the oligonucleotides that encode the binding

region of the peptide. Second, the oligonucleotides can be transferred into another vector, e.g., from phage to phagemid or to p340-1D or to pLamB plasmid. The newly expressed fusion proteins should acquire the same binding activity if the domain is necessary and sufficient for binding to the ligand. This last approach also assesses whether or not flanking amino acid residues encoded by the original vector (i.e., fusion partner) influence peptide binding in any fashion. Third, the oligonucleotides can be synthesized, based on the nucleotide sequence determined for the syngene in the library that encodes the binding peptide, amplified by cloning or PCR amplification using internal and flanking primers cleaved into two pieces and cloned as two half-syngene fragments. In the foregoing manner, the inserted oligonucleotides are subdivided into two equal halves. If the peptide domain important for binding is small, then one recombinant clone would demonstrate binding and the other would not. If neither have binding, then either both are important or the essential portion of the domain spans the middle (which can be tested by expressing just the central region).

Alternatively, by synthesizing peptides corresponding to the deduced sequence of the binding peptide, the binding domains can be analyzed. First, the entire peptide should be synthesized and assessed for binding to the target ligand to verify that the peptide is necessary and sufficient for binding. Second, short peptide fragments, for example, overlapping 10-mers, can be synthesized, based on the amino acid sequence of the random peptide binding domain, and tested to identify those binding the ligand.

In addition, in certain instances, linear motifs may become apparent after comparing the primary structures of different binding peptides from the

library having binding affinity for a target ligand. The contribution of these motifs to binding can be verified with synthesized peptides in competition experiments (i.e., determine the concentration of peptide capable of inhibiting 50% of the binding of the phage to its target;  $IC_{50}$ ). Conversely, the motif or any region suspected to be important for binding can be removed from or mutated within the DNA encoding the random peptide insert and the altered displayed peptide can be retested for binding.

Furthermore, once the binding domain of a random peptide has been identified, differently displayed binding domains can be created by isolating and fusing the binding domain of one random peptide to a new effector domain. The biologically or chemically active effector domain of the peptide can thus be varied. Alternatively, the binding characteristics of an individual peptide can be modified by varying the binding domain sequence to produce a related family of peptides with differing properties for a specific ligand.

Moreover, in a method of directed evolution, the identified random peptides can be improved by additional rounds of random mutagenesis, selection, and amplification of the nucleotide sequences encoding the binding domains. Mutagenesis can be accomplished by creating and cloning a new set of oligonucleotides that differ slightly from the parent sequence, e.g., by 1-10%. Selection and amplification are achieved as described above. By way of example, to verify that the isolated peptides have improved binding characteristics, mutants and the parent phage, differing in their lacZ expression, can be processed together during the screening experiments. Alteration of the original blue-white color ratios during the course of the screening experiment will serve as a visual means to assess the successful selection of

enhanced binders. This process can go through numerous cycles.

#### 5.4. STRUCTURE OF SYNGENES

5           The syngenes of the present invention are synthetic genes, that is, genes that are not derived from a naturally occurring genome. In one embodiment, syngenes are made up of, at least in part, combinations of genes encoding functional domains, such combinations not occurring in nature. Syngenes may be composed of totally synthetic gene sequences or combinations of natural and totally synthetic gene sequences. Syngenes encode a syngene product which is a protein having at least one functional domain. The functional domain is a binding domain with affinity for a ligand. In a preferred aspect, this binding domain is characterized by 1) its strength of binding under specific conditions, 2) the stability of its binding under specific conditions, and 3) its selective specificity for the chosen ligand. In addition the syngene may encode additional functional domains, for example an amino acid sequence that directs cellular trafficking of the syngene product to the desired compartment of the cell; e.g., the tetrapeptide KDEL (SEQ ID NO: 116), KKXX (SEQ ID NO: 118) or KKKXX (SEQ ID NO: 119) where X is any amino acid), to target peptides to the lumen of the endoplasmic reticulum; or, e.g., the tetrapeptide YQRL (SEQ ID NO: 117) to target membrane peptides or secreted peptides to the trans Golgi network (see Luzio and Banting, 1993, TIBS 18:395-398); or a mitochondrial targeting pro-sequence (Singh et al., 1990, Biochem. Biophys. Res. Commun. 169:391-396). A third functional domain may be one that enhances activity of the syngene product, for example a transcriptional activation sequence.

The syngenes thus optionally encode additional sequences besides the encoded binding peptide. Such additional sequences can be markers, linkers, transcriptional activation signals, intracellular localization signals, *in vivo* targeting sequences, processing signals, post-translational modification signals, and the like. The syngene preferably includes regulatory sequences that control the expression of the coding region of the syngene, nucleic acid coding sequences encoding the binding peptide, and optional sequences.

In a preferred embodiment, the syngenes of the invention are created by inserting nucleotides encoding synthetic binding domains discovered by screening libraries as described herein into the context of additional protein coding sequences. This can be done by replacing the DNA binding domain of a known protein or by assembling a totally synthetic gene. In addition to other protein coding sequences, the syngene will generally contain other non-protein coding sequences. These non-protein coding sequences will generally be involved in the control of the expression of the syngene. At a minimum, the syngene preferably contains sequences for initiating transcription and translation, as well as transcriptional processing signals such as a poly A addition signal. The syngene preferably also contains a translation termination signal. Signals for splicing of the RNA transcript encoded by the syngene may also be desirable. The addition of nucleotide sequences in the syngene other than those encoding the binding domain can provide for *in vivo* or intracellular targeting of the syngene or its encoded product or transcriptional and/or post-translational processing events. Thus, in a specific embodiment, the syngene comprises a promoter operably linked to the coding sequences, followed by (3') transcriptional



termination and poly A addition signals. Splice donor and acceptor sequences and sequences affording stability to the construct can also be added. According to one embodiment of the invention, the encoded syngene product can contain an additional region, i.e., an amino acid sequence that functions as a linker domain between one or more of the other domains of the syngene. The presence or absence of the linker domain is optional, as is the type of linker that may be used.

The syngene is preferably incorporated into a vector, for replication, and/or expression of the syngene. Any vector known in the art can be used. In addition to comprising the syngene, the vector comprises appropriate origin(s) of replication, and, preferably, a selectable marker allowing selection of cells expressing the syngene product.

In a specific embodiment, syngene products are hybrid proteins. The amino terminus will generally consist of a methionine residue followed by a spacer sequence of 1-15 amino acids, followed by one or more binding domains (i.e., the sequences identified by screening the peptide library against the ligand of choice). The middle domain of the hybrid protein will often consist of a marker that aids in pre-clinical characterization, e.g. monitoring expression of the syngene *in vitro*, as in, for example, tissue culture cells. This is likely to be a readily detectable epitope or other label, e.g.,  $\beta$ -galactosidase, luciferase or chloramphenicol acetyl transferase genes. This marker generally will be deleted when the syngene is used clinically. A cellular trafficking domain (e.g., KDEL (SEQ ID NO: 116) or YQRL (SEQ ID NO: 117) may next appear; a nuclear localization signal (NLS) at the carboxy terminus is also preferably part of the hybrid protein when it is desired that the syngene product be

targeted to the nucleus. Additional functional domains, e.g., transcriptional activation sequences, or spacers can also be included. Of course, it will be readily apparent to one skilled in the art that the order in which the binding domains, the marker, the trafficking domain, and additional functional domains appear in the syngene may not be critical. The order disclosed above is merely illustrative of one particular possible order. It will also be apparent that, except for the ligand binding domain, the presence of any other domains in a syngene product is optional. An exemplary mRNA expressed by a syngene is shown in Figure 2. The amino terminal sequence is preferably in the range of 3-10 amino acids. The binding domain mediates binding to the ligand of choice. The "middle domain" can be a cellular trafficking signal such as KDEL (SEQ ID NO: 116) or YQRL (SEQ ID NO: 117), or is a marker (label) domain. The additional functional domains (optional) can be transcriptional activation sequences, NLSs, or linkers.

If the syngene product is designed to activate the transcription of a gene, then the syngene will, in addition to the domains discussed above (or instead of the marker domain), contain a transcriptional activation domain. Examples of such domains are: a homopolymeric stretch of glutamine residues, an acidic activation domain (sometimes known as a "negative noodle"), a proline-rich domain, or a serine-threonine-rich domain. It has been demonstrated that when such activation domains are directed to the DNA target, they lead to transcriptional activation of that target (Gerber et al., 1994, Science 263:808-811; Mitchell and Tjian, 1989, Science 245:371-378; Sadowski et al., 1988, Nature 335:563-564; Ptashne and Gann, 1990, Nature 346:329-331).

A nuclear localization signal, such as that of the nucleoplasmin gene, may be added. The nucleoplasmin nuclear localization signal has been shown to function in a variety of contexts, where it leads to the nuclear localization of proteins in which it appears. Nuclear localization signals that can be used in the present invention are not limited to those specifically disclosed herein; they include any that are known in the art. In a particular embodiment, the amino acid coding sequence of a syngene product begins with an amino-terminal spacer sequence, for example, MGASGAS (SEQ ID NO: 120), followed by a nuclear localization sequence which may be KRPAATKKAGQAKKKKR (SEQ ID NO: 121) (the nucleoplasmin NLS, see Kang et al., 1994, Proc. Natl. Acad. Sci. USA 91:340-344) followed by a second spacer sequence, for example, GASGAS (SEQ ID NO: 122), followed by the binding domain. In another embodiment, the amino acid sequences of a syngene product begins with an amino-terminal spacer sequence, followed by the binding domain, followed by a second spacer sequence, followed by the carboxy-terminal nuclear localization signal MISESLRKAIGKR (SEQ ID NO: 123) (Shieh et al., 1993, Plant-Physiol. 101:353-61). In another embodiment, the NLS of the p50 subunit of NF- $\kappa$ B, VQRLRQLM (SEQ ID NO: 124) is used (Henkel et al., 1992, Cell 68:1121-1133). Other nuclear localization signals which can be used as part of a syngene product include but are not limited to those found in various steroid hormone receptors, the SV40 large T antigen, or the consensus sequence therefor, as set forth in Table 7 below.

Table 7

Comparison of Amino Acid Sequences of (Potential) Nuclear Localization Signals in Various Steroid Hormone Receptors and the SV40 Large T Antigen\*

5	Protein	Species	Sequence	SEQ ID NO:
	SV40 large T antigen		126 P K K K R K V P K K K R K V	125 126
10	Progesterone receptor	rabbit human chicken	638(+10) R K F K K F N K 637(+10) ..... 491(+10) ..... L ..	127 128
	Glucocorticoid receptor	human mouse rat	491(+10) R K T K K K I K 498(+10) ..... 510(+10) .....	129
15	Androgen receptor	human rat	628(+10) R K L K K L G N 612(+10) .....	130
	Mineralocorticoid receptor	human	673(+10) R K S K K L G K	131
	Estrogen receptor	human rat mouse chicken Xenopus	256(+11) R K D R R G G R 261(+11) ..... 260(+11) ..... 250(+11) ..... E 251(+11) .....	132 133
20				
	Consensus sequence		R R R R K D . . G G . K K K	134

25 \* Taken from Gulochon-Maniel et al., 1989, Cell 57:1147-1154. The numbers refer to the amino acid position of the first amino acid. In the case of steroid receptors, the distance of this amino acid from the last conserved cysteine in the DNA binding region is indicated in brackets.

30 According to specific embodiments of the invention, the syngene can encode multiple copies of one or more binding domains or additionally contain multiple non-binding domains.

35 Additionally, the syngene can be modified at the base moiety, sugar moiety, or phosphate backbone, to stabilize the syngene, promote *in vivo* transport or localization, etc. Such modifications to bases, sugar

- moieties, and phosphate backbones are made in such a way that the modifications do not interfere with the transcription of the syngene, and thus do not interfere with the expression of the syngene product.
- 5 In a preferred aspect, such modifications are in a noncoding region or at a 5' or 3' terminus of the syngene. For example, the syngene may include other appended groups such as peptides, or agents facilitating transport across the cell membrane (see,
- 10 e.g., Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. WO 88/09810, published December 15, 1988) or blood-brain barrier (see, e.g., PCT Publication No.
- 15 WO 89/10134, published April 25, 1988), hybridization-triggered cleavage agents (see, e.g., Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents (see, e.g., Zon, 1988, Pharm. Res. 5:539-549).
- The syngene may comprise at least one
- 20 modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil,
- 25 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine,
- 30 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil,
- 35 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil,

4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

5 In another embodiment, the syngene comprises at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the syngene  
10 comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl  
15 phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, a portion of the syngene (preferably noncoding) consists of  $\alpha$ -anomeric nucleotides. An  $\alpha$ -anomeric oligonucleotide forms specific double-stranded hybrids with complementary  
20 RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641).

Syngenes of the invention may be synthesized by standard methods known in the art, e.g. by use of  
25 an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), methylphosphonate  
30 oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc. Alternatively, the syngenes are made by recombinant DNA techniques commonly known in the art, e.g. by  
35 replicating a vector comprising the syngene in a suitable host cell.

In a specific embodiment, the syngene comprises a 2'-O-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

#### 5.5. EXPRESSION OF SYNGENES

It will be apparent to one skilled in the art that many satisfactory gene arrangements can be made such that the binding domains of the syngenes of the present invention can be expressed in the desired cell type in vitro or in vivo (see generally, Ausubel et al. (eds.), 1993, Current Protocols in Molecular Biology, John Wiley & Sons, Inc., N.Y.; Kriegler, 1990, Gene Transfer and Expression, A Laboratory Manual, Stockton Press, N.Y.). In a specific embodiment, the syngene is constructed as part of an expression vector that can be introduced in vivo such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, leading to the production of the encoded syngene product. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the syngene product. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the syngene product can be by any promoter known in the art to act in mammalian, preferably human, cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, Cell 22:787-797), the herpes or HIV thymidine kinase

promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42), cytomegalovirus (CMV) promoter, actin  
5 promoter, phosphoglycerate kinase promoter, etc. For example, promoters that function in a wide variety of cell types can be used to obtain expression in a wide variety of cell types where the syngene is introduced. This may be done by using a viral promoter such as a  
10 human cytomegalovirus early gene promoter or the adenovirus major late promoter to drive the expression of the syngene. Such viral promoters are active in a wide variety of mammalian cell types. Alternatively, it is possible to use a construct whereby the syngene  
15 is transcribed off of a constitutive mammalian cell promoter. Examples of such constitutive promoters are actin promoters and the PGK (phosphoglycerol kinase) promoter. Since the latter promoter is functional for high level transcription in all living mammalian cells  
20 it is an especially preferred choice.

In another embodiment the syngene is constructed such that it can be expressed specifically or substantially specifically in a specific tissue or specific cell type. Numerous tissue and cell specific  
25 promoters and control regions have been characterized and it will be apparent to one skilled in the art how to select the appropriate one for expression in the desired cell type. For example, the following tissue specific control regions may be used: elastase I gene  
30 control region which is active in pancreatic acinar cells (Swift et al., 1984, Cell 38:639-646; Ornitz et al., 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in  
35 pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., 1984,



Cell 38:647-658; Adames et al., 1985, Nature 318:533-538; Alexander et al., 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast  
5 cells (Leder et al., 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert et al., 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf et al., 1985, Mol. Cell. Biol. 5:1639-1648; Hammer et al., 1987, Science 235:53-58; alpha 1-  
10 antitrypsin gene control region which is active in the liver (Kelsey et al., 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram et al., 1985, Nature 315:338-340; Kollias et al., 1986, Cell 46:89-94; myelin basic  
15 protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle  
20 (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., 1986, Science 234:1372-1378).

It will be apparent to one skilled in the  
25 art that numerous diverse vectors can be used for the delivery and subsequent expression of syngenes within mammalian cells. Simple shuttle vectors may be used such that a plasmid is grown in *E. coli* and purified  
DNA transferred into mammalian cells by the use of  
30 electroporation, calcium phosphate precipitates, DEAE dextran or with the assistance of liposomes. Such an exemplary shuttle vector is shown in Figure 3. Furthermore many excellent viral based vectors can be used depending upon the therapeutic application. For  
35 instance herpes simplex virus may be used to deliver genes to the brain (Leib and Olivo, 1993, Bioessays 15:547-54), retroviruses (Salmons and Gunzburg, 1993,

Human Gene Therapy 4:129-142) and adeno-associated virus (Walsh et al., 1993, Proc. Soc. Exp. Biol. Med. 204:289-300), which have the advantage of becoming integrated into the cellular DNA, may be used to  
5 deliver syngenes via transduction into hematopoietic, pluripotent stem cells and other somatic cells.

One of skill in the art would recognize that the underlying DNA sequences of the protein coding regions of syngenes may be varied so as to take  
10 advantage of known codon usage frequencies in different organisms. The degeneracy of the genetic code allows for a plurality of DNA sequences to be constructed that all code for the same peptide or protein in a syngene. In some cases, for example, it  
15 may be desirable to introduce changes in the DNA sequence of the syngene encoding the binding domain in order to increase the expression of that binding domain in the cell in which the syngene is to be used. To do this, one would consult any well known table of  
20 codon frequency usage, such as that found in Ausubel et al. (eds.), 1993, Current Protocols in Molecular Biology, John Wiley & Sons, Inc., N.Y., at Appendix 1 A.1.6. Using the information found there, one would use any well known method in the art to change the  
25 codons of the binding domain of the syngene that happen to be little used in the organism in which the syngene is to be expressed into codons that are more common in that organism.

30 5.6. ASSAY FOR THE ABILITY OF A  
SYNGENE PRODUCT TO ACTIVATE  
OR INHIBIT TRANSCRIPTION

Syngenes that encode binding domains that bind to a DNA transcriptional regulatory region, but that do not contain a transcriptional activation  
35 signal, can be tested for their ability to inhibit the transcription of a gene by the following assay,

presented by way of example but not limitation. The assay is carried out by using three sets of plasmids in an established cell line. The cell line expresses a transcription factor that has a DNA binding site  
5 that is the same as the DNA binding site of the syngene product that is to be assayed. The three plasmids have the following characteristics:

Plasmid 1: This plasmid contains the syngene to be tested cloned into a vector that will direct the  
10 expression of the syngene in the cell line;

Plasmid 2: This plasmid contains a reporter gene. The promoter of the reporter gene contains the DNA sequence which forms the binding site for the transcription factor expressed by the cells and for  
15 the binding domain encoded by the syngene. Normally, this DNA sequence will be essentially the same sequence that was used to screen the library in the process of obtaining the binding domain encoded by the syngene. The reporter gene codes for the expression  
20 of any detectable marker such as, for example, chloramphenicol acetyl transferase (CAT),  $\beta$ -galactosidase, horse radish peroxidase, etc., under the control of the promoter.

Plasmid 3: This is a control plasmid to  
25 monitor the efficiency of transfection. It directs the expression of a second reporter gene expressing a product that can be easily assayed. The activity of the promoter that drives the expression of this product is not dependent upon binding of a  
30 transcription factor to the binding site of the syngene encoded binding domain that is being assayed. Preferably, the promoter is constitutive in the cell being used. Alternatively, it is an inducible promoter, and the assays described below are carried  
35 out in the presence of its inducer. The reporter gene in plasmid 3 is different from that in plasmid 2.

Two procedures are done, consisting of a test procedure and a control procedure. In the test procedure, the three plasmids are simultaneously co-introduced into each cell. Introduction can be by any  
5 of the well known methods in the art. Such methods include electroporation, calcium phosphate mediated transfection, and liposome mediated transfection. Preferably, introduction is by electroporation. In the control procedure, plasmids 2 and 3 are co-  
10 introduced, without plasmid 1 (which contains the syngene).

After culture for a time sufficient to allow expression of the products encoded by the plasmids, the amounts of expressed reporter gene products from  
15 plasmids 2 and 3 are measured. If the syngene product is able to compete with the transcription factor for binding to the DNA binding sequence in the promoter of the first reporter gene on plasmid 2, then the level of expression of the first reporter gene product will  
20 be low. This is because binding of the syngene product to the DNA binding sequence in the promoter of the first reporter gene will prevent binding of the transcription factor to that sequence. Therefore, the transcription factor will not be able to activate the  
25 transcription of the first reporter gene. If the syngene product is not able to compete with the transcription factor, then the level of expression of the first reporter gene product will be high. This will be the case in the control procedure, where  
30 plasmid 1, which contains the syngene, is not used.

The level of expression of the first reporter gene product (from plasmid 2) is compared to the level of expression of the second reporter gene product (from plasmid 3) to determine the ratio of  
35 expression of the two products. This ratio is determined for both the test procedure with, and the control procedure without, plasmid 1 containing the

syngene. When the ratio from the test procedure is less than the ratio from the control procedure, this indicates that the syngene product is competing with the transcription factor for binding to the DNA binding sequence in the promoter of the first reporter gene. When the ratio from the test procedure is equal to the ratio from the control procedure, this indicates that the syngene product is unable to compete with the transcription factor.

By way of example, to determine if a syngene product can inhibit the transcription of a gene the transcription of which depends upon binding of the transcription factor NF- $\kappa$ B to its DNA binding site, a human cell line that expresses NF- $\kappa$ B (p65) is used (e.g., Jurkat, a T cell line; Okamoto et al., 1994, J. Biol. Chem. 269:8582-8587). The first plasmid contains the syngene which is driven by the adenovirus major late promoter or by some other promoter that is active in the cells being used. The second plasmid contains the first reporter gene which consists of the H2kb promoter (Baldwin and Sharp, 1988, Proc. Natl. Acad. Sci. USA 85:723-727) directing the expression of the CAT gene. The H2kb promoter is dependent upon NF- $\kappa$ B binding for its activity. The control plasmid consists of the luciferase gene (the second reporter gene) driven by a promoter that does not depend upon NF- $\kappa$ B for its expression.

The three plasmids are introduced into the cells via electroporation. When the syngene that is being assayed expresses a binding domain that is specific for the NF- $\kappa$ B binding site of the H2kb gene but does not also contain a transcriptional activation site, then the ratio of CAT activity to luciferase activity will be low when compared to the same ratio obtained when no syngene, or a control syngene that does not encode a binding site that is specific for the NF- $\kappa$ B binding site, is used.

This is because the syngene that encodes a binding domain that specifically binds the NF- $\kappa$ B binding site will bind the NF- $\kappa$ B site in the reporter plasmid containing the CAT gene. This will prevent  
5 binding of the cell's endogenous NF- $\kappa$ B transcription factor to that site, thus preventing NF- $\kappa$ B from activating the promoter of the reporter gene.

To determine whether a syngene-encoded protein product can activate the transcription of a  
10 gene, the same three plasmids can be used as are described above in the test and control procedures. The cell line used in the assay for activation differs, however. To test for activation, a cell line must be used that does not express the transcription  
15 factor whose binding site is the same as that of the syngene product. Because the cell line does not express the transcription factor, the first reporter gene on the second plasmid will not be transcribed in the absence of a syngene product. If the syngene  
20 product is capable of binding to the DNA binding sequence in the promoter of the first reporter gene and is also capable of activating transcription, the reporter gene on the second plasmid will be expressed. In this assay, the syngene to be tested will  
25 preferably encode a transcriptional activation signal in addition to a binding domain.

Once again, the ratio of first reporter gene product over second reporter gene products is compared for the test and control procedures (with and without  
30 the syngene). In this case, however, the presence in the test procedure of a syngene that encodes a product that is capable of binding to the promoter of, and activating the transcription of, the first reporter gene will result in a larger value for the ratio in  
35 the case of the test procedure as compared to the control procedure. The ratio for the control procedure is smaller, reflecting the inactivity of the

promoter of the first reporter gene in the absence of the transcription factor and the syngene product.

As a specific example, to determine whether a syngene can activate the transcription of a gene the transcription of which depends upon the binding of the transcription factor NF- $\kappa$ B to its DNA binding site, a cell line is used that substantially does not express NF- $\kappa$ B. HeLa cells (Scheinman et al., 1993, Mol. Cell. Biol. 13:6089-6101) express only low levels of endogenous NF- $\kappa$ B and thus can be used. Alternatively, F9 embryonal carcinoma cells can be used. The same three plasmids as above are co-transfected into this cell line. If the syngene product contains a binding domain specific for the NF- $\kappa$ B site as well as a transcriptional activation domain, the ratio of CAT activity to luciferase activity will be high when compared to the same ratio obtained when no syngene, or a control syngene product that lacks a binding domain specific for the NF- $\kappa$ B site, is used. When no syngene, or a control syngene, is used, CAT activity will be very low because the cell, lacking NF- $\kappa$ B activity, will transcribe the H2 $\kappa$ b promoter (which drives expression of the CAT gene) at a very low rate. Thus, there will be very little CAT activity unless the syngene to be assayed binds to the NF- $\kappa$ B site on the H2 $\kappa$ b promoter and activates expression of the CAT gene.

If a cell line that does not express a transcription factor of interest is not readily available, it may be made by methods known in the art. For example, the gene for the transcription factor of interest may be inactivated by promoting homologous recombination of that gene with a non-functional copy of itself (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

A preferred vector for expression of a syngene for characterization of its biological activity is that described by Pei et al., 1991, J. Biol. Chem. 266:9598-9604. This vector consists of an  
5 adenovirus major late promoter driving a gene cassette with the adenovirus tripartite leader sequence and a hemoglobin splice donor and acceptor followed by the desired coding sequences followed by the  
10 polyadenylation site and 3' non-coding sequences from the SV40 virus early transcriptional unit. In addition, this vector has the pBR322 replicon and an antibiotic resistance marker for growth in *E. coli*.

Although the above discussion was couched in terms of NF- $\kappa$ B binding sites, it will be readily  
15 understood by one of skill in the art that the same general method of testing the ability of a syngene to activate or inhibit transcription can be extended to test the activity of syngene products that bind to various other transcription factor binding sites, or  
20 that bind to a ligand of choice and thereby activate or inhibit an activity that is detectable in the assay.

#### 25 5.7. DIFFERENTIAL REGULATION OF CLOSELY RELATED TRANSCRIPTIONAL REGULATORY SITES

It is well known that the DNA sequences that define the binding sites of many transcription factors differ somewhat from gene to gene. Thus, the NF- $\kappa$ B binding site in the HLA class I genes differs from the  
30 NF- $\kappa$ B site in the IL-6 gene or the IL-8 gene (respectively termed IL-6 $\kappa$ B and IL-8 $\kappa$ B). See Section 6.1.1. While, for some transcription factors, slightly different sites are recognized by slightly different versions of the transcription factors, in  
35 some cases the same transcription factor recognizes the somewhat different versions of its binding site.



In those cases where the same transcription factor recognizes different binding sites, the use of the syngenes of the present invention affords the opportunity for achieving a level of regulation of the different sites that surpasses that which could be achieved by the use of the natural transcription factor. This is because it is possible to isolate highly specific DNA binding domains from random synthetic peptide libraries according to the present invention that bind specifically to one member of the related binding sites of a given transcription factor but that do not bind to the other sites. This is done by screening a random synthetic peptide library with a first oligonucleotide ligand that represents the DNA binding site that it is desired to specifically regulate. After isolating a group of binding domains that bind to this first oligonucleotide, one then makes use of other oligonucleotides that are closely related, but slightly different from the first oligonucleotide. By selecting against binding domains that will bind these other oligonucleotides (by use of the methods described in Section 6.1 and its subsections), one obtains a highly specific DNA binding domain, i.e. a binding domain that specifically binds only to the first oligonucleotide, and not to the others. Using this approach, for example, it is possible to regulate the activity of the IL-6 gene (by use of a syngene expressing an HSDB specific to the NF- $\kappa$ B site of the IL-6 gene) without affecting the activity of the IL-8 gene.

#### 5.8. USES OF SYNGENES

The invention also provides methods for use of the syngenes, e.g., in diagnosis and therapy of various disorders.

The present invention vastly expands the number of genes that are available for use in gene

therapy. The present invention provides methods for finding synthetic nucleic acids encoding proteins with diagnostic or therapeutic value, or value in in vitro assays. The present invention further provides  
5 methods for delivery of such peptides or proteins in vivo by expression in vivo from the administered syngene.

The present invention provides methods for using synthetic peptides or proteins, and the genes  
10 encoding them, to fulfill roles that naturally occurring proteins have not necessarily been evolved for. One need only identify a target whose activity it would be desirable to modulate. Given such a target, a synthetic peptide in a random peptide  
15 library that binds the target and can thus modulate the target's activity can be readily identified by the present invention. The present invention also provides for the identification of a nucleic acid (syngene) that encodes the peptide. Such a syngene  
20 can be used, e.g., in gene therapy. As an example, the present invention provides methods for identifying syngenes that inhibit or enhance the transcriptional activity of a wide variety of naturally occurring genes, preferably with a specificity not found in  
25 natural systems. In a specific embodiment, syngenes can be used to effect differential regulation of closely related transcriptional regulatory sites as described in Section 5.7.

In addition to affecting transcription,  
30 among the other nuclear processes that syngenes may be useful in modulating are: post-transcriptional processing of RNA, DNA repair, and DNA replication. Syngene products may bind to actin and thus be of use in regulating mitotic spindle formation and cell  
35 division.

Syngenes are also useful for modulating processes that occur outside of the nucleus. For

example, in the cytoplasm, the encoded protein of a syngene, produced by intracellular expression of the syngene, may be used to affect the activity of inhibitors of transcription factors such as IF- $\kappa$ B.

- 5 They may also be used to modulate signal transduction pathways, metabolic pathways, RNA translation, and intracellular trafficking. In the cell membrane, syngenes may be used to modulate the activity of membrane receptors, ion channels, or exocytotic and  
10 endocytotic pathways.

- In tissue, syngenes, via expression of their encoded proteins, may be used to regulate cell/cell signalling and transcytosis. Cell/cell junctions and the extracellular matrix are appropriate targets for  
15 syngenes. Syngenes may be used to regulate cell adhesion or cell/cell recognition. In the general circulation, syngenes may be used to regulate the activity of receptor ligands.

- Syngenes can be used in any appropriate  
20 method of gene therapy, as would be recognized by those in the art upon considering this disclosure. The resulting action of a syngene in the gene therapy patient can, for example, lead to the activation or inhibition of a preselected gene in the patient, thus  
25 leading to improvement of the diseased condition afflicting the patient. Methods of gene therapy are detailed in Section 5.8.1.

- Alternatively, the syngenes may be introduced into appropriate host cells and thereby  
30 used for the recombinant production of their encoded peptides. The peptides thus produced can be used, e.g., in in vitro assays to detect and/or quantitate the ligand of choice to which they bind, for therapeutic purposes, diagnostic purposes, or to  
35 regulate transcription (e.g., where the ligand of choice modulates transcription).

In another aspect, the invention relates to the peptide products of syngenes and their therapeutic and diagnostic uses. The syngene-encoded peptides can be used therapeutically and diagnostically, to  
5 regulate transcription, signal transduction, etc. as described herein for syngenes. The syngene products can be used in in vitro binding assays, to detect and/or measure amounts of their binding partners (e.g., the ligand of choice). The syngene products  
10 can be used in vivo, e.g., labeled with an appropriate marker, to image their binding partners.

#### 5.8.1. GENE THERAPY

In its broadest sense, gene therapy refers  
15 to therapy performed by the administration of a nucleic acid to a subject. The nucleic acid, either directly or indirectly via its encoded protein, mediates a therapeutic effect in the subject.

The syngenes of the present invention can be  
20 used in any of the methods for gene therapy available in the art. The descriptions below are meant to be illustrative of such methods. It will be readily understood by those of skill in the art that the methods illustrated represent only a sample of all  
25 available methods of gene therapy.

For general reviews of the methods of gene therapy, see Goldspiel et al., 1993, Clinical Pharmacy 12:488-505; Wu and Wu, 1991, Biotherapy 3:87-95; Tolstoshev, 1993, Ann. Rev. Pharmacol. Toxicol.  
30 32:573-596; Mulligan, 1993, Science 260:926-932; and Morgan and Anderson, 1993, Ann. Rev. Biochem. 62:191-217; May, 1993, TIBTECH 11(5):155-215). Methods commonly known in the art of recombinant DNA technology which can be used are described in Ausubel  
35 et al. (eds.), 1993, Current Protocols in Molecular Biology, John Wiley & Sons, NY; and Kriegler, 1990,

Gene Transfer and Expression, A Laboratory Manual,  
Stockton Press, NY.

Delivery of the syngene into a patient may  
be either direct, in which case the patient is  
5 directly exposed to the syngene or syngene-carrying  
vector, or indirect, in which case, cells are first  
transformed with the syngene *in vitro*, then  
transplanted into the patient. These two approaches  
are known, respectively, as *in vivo* or *ex vivo* gene  
10 therapy.

In a specific embodiment, the syngene is  
directly administered *in vivo* for therapeutic effect,  
whereby it is expressed to produce the syngene  
product. This can be accomplished by any of numerous  
15 methods known in the art, e.g., by constructing it as  
part of an appropriate nucleic acid expression vector  
and administering it so that it becomes intracellular,  
e.g., by infection using a defective or attenuated  
retroviral or other viral vector (see U.S. Patent No.  
20 4,980,286), or by direct injection of naked DNA, or by  
use of microparticle bombardment (e.g., a gene gun;  
Biolistic, Dupont), or coating with lipids or cell-  
surface receptors or transfecting agents,  
encapsulation in liposomes, microparticles, or  
25 microcapsules, or by administering it in linkage to a  
peptide which is known to enter the nucleus, by  
administering it in linkage to a ligand subject to  
receptor-mediated endocytosis (see e.g., Wu and Wu,  
1987, J. Biol. Chem. 262:4429-4432) (which can be used  
30 to target cell types specifically expressing the  
receptors), etc. In another embodiment, a syngene-  
ligand complex can be formed in which the ligand  
comprises a fusogenic viral peptide to disrupt  
endosomes, allowing the syngene to avoid lysosomal  
35 degradation. In yet another embodiment, the syngene  
can be targeted *in vivo* for cell specific uptake and  
expression, by targeting a specific receptor (see,

e.g., PCT Publications WO 92/06180 dated April 16, 1992 (Wu et al.); WO 92/22635 dated December 23, 1992 (Wilson et al.); WO92/20316 dated November 26, 1992 (Findeis et al.); WO93/14188 dated July 22, 1993 (Clarke et al.), WO 93/20221 dated October 14, 1993 (Young)). Alternatively, the syngene can be introduced intracellularly and incorporated within host cell DNA for expression, by homologous recombination (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

One common method of practicing gene therapy is by making use of retroviral vectors (see Miller et al., 1993, Meth. Enzymol. 217:581-599). A retroviral vector is a retrovirus that has been modified to incorporate a preselected gene in order to effect the expression of that gene. It has been found that many of the naturally occurring DNA sequences of retroviruses are dispensable in retroviral vectors. Only a small subset of the naturally occurring DNA sequences of retroviruses is necessary. In general, a retroviral vector must contain all of the *cis*-acting sequences necessary for the packaging and integration of the viral genome. These *cis*-acting sequences are:

- a) a long terminal repeat (LTR), or portions thereof, at each end of the vector;
- b) primer binding sites for negative and positive strand DNA synthesis; and
- c) a packaging signal, necessary for the incorporation of genomic RNA into virions.

The gene to be used in gene therapy is cloned into the vector, which facilitates delivery of the gene into a patient.

More detail about retroviral vectors can be found in Boesen et al., 1994, Biotherapy 6:291-302, which describes the use of a retroviral vector to deliver the *mdr1* gene to hematopoietic stem cells in

order to make the stem cells more resistant to chemotherapy. Other references illustrating the use of retroviral vectors in gene therapy are: Clowes et al., 1994, J. Clin. Invest. 93:644-651; Kiem et al., 5 1994, Blood 83:1467-1473; Salmons and Gunzberg, 1993, Human Gene Therapy 4:129-141; and Grossman and Wilson, 1993, Curr. Opin. in Genetics and Devel. 3:110-114.

Adenoviruses are also of use in gene therapy. Adenoviruses are especially attractive 10 vehicles for delivering genes to respiratory epithelia. Adenoviruses naturally infect respiratory epithelia where they cause a mild disease. Other targets for adenovirus-based delivery systems are liver, the central nervous system, endothelial cells, 15 and muscle. Adenoviruses have the advantage of being capable of infecting non-dividing cells. Kozarsky and Wilson, 1993, Current Opinion in Genetics and Development 3:499-503 present a review of adenovirus-based gene therapy. Bout et al., 1994, Human Gene 20 Therapy 5:3-10 demonstrated the use of adenovirus vectors to transfer genes to the respiratory epithelia of rhesus monkeys. Other instances of the use of adenoviruses in gene therapy can be found in Rosenfeld et al., 1991, Science 252:431-434; Rosenfeld et al., 25 1992, Cell 68:143-155; and Mastrangeli et al., 1993, J. Clin. Invest. 91:225-234.

It has been proposed that adeno-associated virus (AAV) be used in gene therapy (Walsh et al., 1993, Proc. Soc. Exp. Biol. Med. 204:289-300).

30 Another approach to gene therapy involves transferring a gene to cells in tissue culture by such methods as electroporation, lipofection, calcium phosphate mediated transfection, or viral infection. Usually, the method of transfer includes the transfer 35 of a selectable marker to the cells. The cells are then placed under selection to isolate those cells

that have taken up and are expressing the transferred gene. Those cells are then delivered to a patient.

In this embodiment, the syngene is introduced into a cell prior to administration *in vivo* of the resulting recombinant cell. Such introduction can be carried out by any method known in the art, including but not limited to transfection, electroporation, microinjection, infection with a viral or bacteriophage vector containing the syngene sequences, cell fusion, chromosome-mediated gene transfer, microcell-mediated gene transfer, spheroplast fusion, etc. Numerous techniques are known in the art for the introduction of foreign genes into cells (see e.g., Loeffler and Behr, 1993, Meth. Enzymol. 217:599-618; Cohen et al., 1993, Meth. Enzymol. 217:618-644; Cline, 1985, Pharmac. Ther. 29:69-92) and may be used in accordance with the present invention, provided that the necessary developmental and physiological functions of the recipient cells are not disrupted. The technique should provide for the stable transfer of the syngene to the cell, so that the syngene is expressible by the cell and preferably heritable and expressible by its cell progeny.

The resulting recombinant cells can be delivered to a patient by various methods known in the art. In a preferred embodiment, epithelial cells are injected, e.g., subcutaneously. In another embodiment, recombinant skin cells may be applied as a skin graft onto the patient. Recombinant blood cells (e.g., hematopoietic stem or progenitor cells) are preferably administered intravenously. The amount of cells envisioned for use depends on the desired effect, patient state, etc., and can be determined by one skilled in the art.

Cells into which a syngene can be introduced for purposes of gene therapy encompass any desired,



available cell type, and include but are not limited to epithelial cells, endothelial cells, keratinocytes, fibroblasts, muscle cells, hepatocytes; blood cells such as T lymphocytes, B lymphocytes, monocytes, macrophages, neutrophils, eosinophils, megakaryocytes, granulocytes; various stem or progenitor cells, in particular hematopoietic stem or progenitor cells, e.g., as obtained from bone marrow, umbilical cord blood, peripheral blood, fetal liver, etc.

10 In a preferred embodiment, the cell used for gene therapy is autologous to the patient.

In an embodiment in which recombinant cells are used in gene therapy, a syngene is introduced into the cells such that it is expressible by the cells or their progeny, and the recombinant cells are then administered *in vivo* for therapeutic effect; stem and progenitor cells are preferred for use. Any stem and/or progenitor cells which can be isolated and maintained *in vitro* can potentially be used in accordance with this embodiment of the present invention. Such stem cells include but are not limited to hematopoietic stem cells (HSC), stem cells of epithelial tissues such as the skin and the lining of the gut, embryonic heart muscle cells, liver stem cells (PCT Publication WO 94/08598, dated April 28, 1994), and neural stem cells (Stemple and Anderson, 1992, Cell 71:973-985).

Epithelial stem cells (ESCs) or keratinocytes can be obtained from tissues such as the skin and the lining of the gut by known procedures (Rheinwald, 1980, Meth. Cell Bio. 21A:229). In stratified epithelial tissue such as the skin, renewal occurs by mitosis of stem cells within the germinal layer, the layer closest to the basal lamina. Stem cells within the lining of the gut provide for a rapid renewal rate of this tissue. ESCs or keratinocytes obtained from the skin or lining of the gut of a

patient or donor can be grown in tissue culture (Rheinwald, 1980, Meth. Cell Bio. 21A:229; Pittelkow and Scott, 1986, Mayo Clinic Proc. 61:771). If the ESCs are provided by a donor, a method for suppression of host versus graft reactivity (e.g., irradiation, drug or antibody administration to promote moderate immunosuppression) can also be used.

With respect to hematopoietic stem cells (HSC), any technique which provides for the isolation, propagation, and maintenance in vitro of HSC can be used in this embodiment of the invention. Techniques by which this may be accomplished include (a) the isolation and establishment of HSC cultures from bone marrow cells isolated from the future host, or a donor, or (b) the use of previously established long-term HSC cultures, which may be allogeneic or xenogeneic. Non-autologous HSC are used preferably in conjunction with a method of suppressing transplantation immune reactions of the future host/patient. In a particular embodiment of the present invention, human bone marrow cells can be obtained from the posterior iliac crest by needle aspiration (see, e.g., Kodo et al., 1984, J. Clin. Invest. 73:1377-1384). In a preferred embodiment of the present invention, the HSCs can be made highly enriched or in substantially pure form. This enrichment can be accomplished before, during, or after long-term culturing, and can be done by any techniques known in the art. Long-term cultures of bone marrow cells can be established and maintained by using, for example, modified Dexter cell culture techniques (Dexter et al., 1977, J. Cell Physiol. 91:335) or Witlock-Witte culture techniques (Witlock and Witte, 1982, Proc. Natl. Acad. Sci. USA 79:3608-3612).

In a specific embodiment, the syngene to be introduced for purposes of gene therapy comprises an

inducible promoter operably linked to the coding region, such that expression of the syngene is controllable by controlling the presence or absence of the appropriate inducer of transcription.

5

#### 5.9. PHARMACEUTICAL COMPOSITIONS

The invention provides methods of treatment by administration to a subject of an effective amount of a pharmaceutical (therapeutic) composition comprising a syngene. Such a syngene envisioned for therapeutic use is referred to hereinafter as a "Therapeutic" or "Therapeutic of the invention." As used hereinbelow, "Therapeutic" or "Therapeutic of the invention" shall also be used to refer to molecules comprising the encoded peptide products of syngenes, e.g., where a molecule comprising the peptide binding domain encoded by a syngene is used therapeutically. In a preferred aspect, the Therapeutic is substantially purified. The subject is preferably an animal, including but not limited to animals such as cows, pigs, horses, chickens, cats, dogs, etc., and is preferably a mammal, and most preferably human.

Formulations and methods of administration that can be employed when the Therapeutic comprises a syngene are described in Section 5.8.1; additional appropriate formulations and routes of administration can be selected from among those described hereinbelow.

Various delivery systems are known and can be used to administer a Therapeutic of the invention, e.g., encapsulation in liposomes, microparticles, microcapsules, recombinant cells containing the Therapeutic, receptor-mediated endocytosis (see, e.g., Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432), construction of a Therapeutic syngene as part of a retroviral or other vector, etc. Methods of introduction include but are not limited to

intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, and oral routes. The compounds may be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically active agents. Administration can be systemic or local. In addition, it may be desirable to introduce the pharmaceutical compositions of the invention into the central nervous system by any suitable route, including intraventricular and intrathecal injection; intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir. In a specific embodiment, it may be desirable to utilize liposomes targeted via antibodies to specific identifiable cell surface antigens.

In a preferred aspect, the Therapeutic comprises a syngene that is part of an expression vector that expresses the syngene in a suitable host. In particular, such a syngene has a promoter operably linked to the syngene coding region, said promoter being inducible or constitutive. In another particular embodiment, the syngene is a nucleic acid molecule in which the syngene coding sequences and any other desired sequences are flanked by regions that promote homologous recombination at a desired site in the genome, thus providing for intrachromosomal expression of the syngene (Koller and Smithies, 1989, Proc. Natl. Acad. Sci. USA 86:8932-8935; Zijlstra et al., 1989, Nature 342:435-438).

In a specific embodiment, it may be desirable to administer the Therapeutics of the invention locally to the area in need of treatment; this may be achieved by, for example, and not by way

of limitation, local infusion during surgery, topical application, e.g., in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an  
5 implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

The present invention provides pharmaceutical compositions. Such compositions  
10 comprise a therapeutically effective amount of a Therapeutic, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The  
15 carrier and composition can be sterile. The formulation should suit the mode of administration.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can  
20 be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include  
25 standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc.

In a preferred embodiment, the composition is formulated in accordance with routine procedures as  
30 a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a  
35 solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either

separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The Therapeutics of the invention can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The amount of the Therapeutic of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, in vitro assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a

notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or  
5 sale for human administration.

## 6. EXAMPLES

For the examples incorporated herein, a TSAR library is utilized; however, to those skilled in the art, it will be apparent that other random peptide display libraries may be used. An example of a TSAR library is the TSAR-9 library disclosed in Kay et al., 1993, Gene 128:59-65. TSAR-9 constructs display a peptide of about 38 amino acids in length having 36  
15 totally random positions.

### 6.1. ISOLATION OF HSDBs FOR NF- $\kappa$ B AND NF-IL6 BINDING SITES

#### 20 6.1.1. OLIGONUCLEOTIDE TARGETS

The synthetic oligonucleotide targets (ligands) for isolating HSDBs that encode peptide sequences that bind to specific cis-elements for several genes of interest are described in the  
25 following paragraphs.

For the NF- $\kappa$ B site of HLA class I genes the sequence is 5'gggtGGGGATTCCCCatct (SEQ ID NO: 135) (called H2 $\kappa$ b-U) and 5'agatGGGGAATCCCCaccc (SEQ ID NO: 136) (called H2 $\kappa$ b-L) (the suffix U stands for upper strand and a suffix L stands for lower strand). The  
30 double stranded oligonucleotide formed by the hybridization of H2 $\kappa$ B-U and H2 $\kappa$ B-L is called the H2 $\kappa$ B oligonucleotide. The upper case sequence is the known NF- $\kappa$ B binding site (see Baldwin and Sharp, 1988, Proc. Natl. Acad. Sci, USA 85:723-727). This particular  
35 sequence is homologous to the NF- $\kappa$ B regulatory domain of the murine homologue of the human HLA class I gene.

The murine sequence is from the H-2 $\kappa$ (d) gene and is in the databases as GenBank accession X01815. The upper case NF- $\kappa$ B site is completely conserved in human HLA class I genes such as HLA-B27 (GenBank M12967) and HLA-J (GenBank M80470) and HLA-A2 (GenBank K02883).

IL-6 is an attractive target for regulation by syngenes since it has been shown to be involved in immune responses and the acute phase protein responses (Kishimoto, 1989, Blood 74:1-10; Clark, 1989, Ann. N.Y. Acad. Sci. 557:438-443; Castell et al., 1989, Ann. N.Y. Acad. Sci. 557:87-101). For the NF- $\kappa$ B site in the IL-6 gene, the following oligonucleotides are used: 5'atgtGGGATTTTCCcatg (SEQ ID NO: 137) (called IL6 $\kappa$ B-U) and 5'catgGGAAAATCCacat (SEQ ID NO: 138) (called IL6 $\kappa$ B-L). The double stranded oligonucleotide formed by the hybridization of IL6 $\kappa$ B-U and IL6 $\kappa$ B-L is called the IL6 $\kappa$ B oligonucleotide. This site has been defined as the NF- $\kappa$ B regulatory sequence in the IL-6 gene by Taiji et al., 1993, Proc. Natl. Acad. Sci. USA 90:10193-10197.

For the NF- $\kappa$ B site in the IL-8 gene, the following oligonucleotides are used: 5'atcgTGGAATTCctctg (SEQ ID NO: 139) (called IL8 $\kappa$ B-U) and 5'cagaGGAAATCCAcgat (SEQ ID NO: 140) (called IL8 $\kappa$ B-L). The double stranded oligonucleotide formed by the hybridization of IL8 $\kappa$ B-U and IL8 $\kappa$ B-L is called the IL8 $\kappa$ B oligonucleotide. This site has been defined as the NF- $\kappa$ B regulatory sequence in the IL-8 gene by Kunsch and Rosen, 1993, Mol. Cell. Biol. 13:6137-6146.

For the NF-IL6 site found in the IL6 gene, the following oligonucleotides are synthesized: 5'acgtcATTGCACAAtcctt (SEQ ID NO: 141) (called NFIL6-U) and 5'aagaTTGTGCAATgacgt (SEQ ID NO: 142) (called NFIL6-L). The double stranded oligonucleotide formed by the hybridization of NFIL6-U and NFIL6-L is called the NFIL6 oligonucleotide. Those nucleotides in upper case correspond to the consensus sequence for an NF-



IL6 site and have been shown to constitute a functional NF-IL6 regulatory cis-element (Zhang et al, 1994, Proc. Natl. Acad. Sci. USA 91:2225-2229).

For the experiments outlined below where it is desirable to immobilize the oligonucleotides to streptavidin coated plastic surfaces, all the upper strand oligonucleotides listed above are synthesized with a biotinylated thymidine nearest to the 3' thymidine position, but not at the 3' end. This biotin moiety is separated from the thymidine base by a 20 carbon spacer arm. Therefore, when the respective upper and lower strand oligonucleotide pairs are annealed together, there is one biotin tag on each pair. By kinasing the 5' end of each synthetic oligonucleotide with T4 polynucleotide kinase as described in Ausubel et al. (eds.), 1993, *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., New York, NY, then annealing the respective pairs of oligonucleotides in 10 mM Tris, 1mM EDTA, pH 7.5 (TE buffer) with 50 mM NaCl at 30°C and then carrying out a ligation reaction (section 3.14.2 in *Current Protocols in Molecular Biology*), one obtains DNA molecules with one or more transcription factor binding sites per molecule. Unless otherwise stated, all abbreviations and buffers are as described in the Appendix in *Current Protocols in Molecular Biology*. Thus, for the isolation of TSAR phage that bind to DNA, one can prepare either a monovalent target, solely by annealing the appropriate pairs of oligonucleotides or a multivalent target by the ligation of kinased monovalent target molecules. The ligated fragments of DNA can be analyzed by gel electrophoresis with visualization of ethidium bromide stained DNA using short wave UV light to determine the average size of a ligated fragment and thus, the average number of cis-elements per fragment. For the experiments carried out below, it is optimal to have

either one site per molecule (unligated fragments) or 10 sites per molecule. Other numbers of sites are less desirable. Fragments of the appropriate size can be recovered by gel electrophoresis or velocity sedimentation centrifugation techniques.

#### 6.1.2. SCREENING

Phage from a TSAR library that bind to specific DNA sequences are isolated by a process of screening called panning. In a preferred method, "Universal Covalent" multiwell plates obtained from Costar (Cambridge, MA), or similar multiwell plates, are utilized. The specific DNA target is immobilized directly to the wells in a procedure supplied by the manufacturer. Costar's Universal Covalent surface is intended to covalently immobilize biomolecules via an abstractable hydrogen using UV illumination resulting in a carbon-carbon bond. Although the linkage is non-specific and does not allow for site-directed orientation of a biomolecule, this surface is very useful for the immobilization of double stranded DNA. A 100  $\mu$ l DNA sample (10  $\mu$ g/ml in 50 mM Na Acetate buffer) is added to each of the appropriate wells and incubated at room temperature for 1 hr. The solution is removed and the DNA covalently immobilized to the surface by exposing the plate to UV light for the time determined to be optimal in the UV Intensity Verification section of the instructions provided by Costar. Subsequently, each well is blocked with BSA as described below. The advantage to this method is that one obtains fewer TSAR binding phage targeted to a substance such as streptavidin which, in an alternative method, is used to attach the biotinylated DNA to the multi-well plate surface. However, a disadvantage is that the DNA is bound directly to the plate, providing some degree of steric interference with the binding of the desired TSAR phage to the DNA

target. This is less of a problem when using target DNA immobilized using streptavidin due to the 20 carbon space linker between the biotin and the DNA.

- Another method by which panning can be carried out is as follows. Aliquots of phage are first placed into ELISA plate wells coated only with streptavidin. This removes those phage capable of binding to streptavidin. Then the aliquots are transferred to wells coated with target DNA immobilized via a biotin linked to streptavidin bound to the surface of an ELISA plate well. Preferably, Reacti-Bind brand streptavidin coated strip plates, cat. # 15124 are used. These are obtainable from Pierce, 3747 North Meridian Road, Rockford, IL.
- Less preferably, one may coat Costar brand E.I.A./R.I.A. multi-well plates (cat. #3590). All wells are coated with streptavidin using 1  $\mu$ g/ml in 100 mM Na Bicarbonate, pH 8.4 buffer. 50  $\mu$ l are added to each well and incubated at 4°C overnight or 22°C for 4 hrs. The streptavidin solution is pipetted away, and the remaining binding surfaces are blocked with 300  $\mu$ l/well of a solution containing 5 mg/ml BSA in 100 mM Na Bicarbonate, pH 8.4 buffer for 1 hr at 37°C or overnight at 4°C.
- Initially, the wells on a coated 96-well multi-well plate are labelled in pairs. Two wells are used for each specific DNA target, one well for each target is labelled "A" and the other is labeled "B". The wells are washed with a Wash Buffer composed of 1X phosphate buffered saline (PBS), 0.1% Tween <sup>®</sup> 20 and 0.1% bovine serum albumin (BSA). The washing is repeated three times.

- The appropriate oligonucleotide fragments (specific target DNA) are added to wells labelled "B" using 50  $\mu$ l of a DNA solution (20  $\mu$ g/ml) in 1x PBS buffer. The streptavidin is allowed to capture the DNA for 2 hrs at room temperature. The DNA solutions

are withdrawn with a pipette and discarded. All non-specifically bound DNA is washed away using wash buffer described above.

To begin the actual panning,  $10^{11}$  particles of phage from a TSAR-9 or TSAR-12 or other suitable library in 100  $\mu$ l in 1X PBS with 0.1% BSA are added to each well labelled "A" and allowed to incubated for 2 hrs at room temperature. To begin the specific adsorption of phage that bind DNA, the adsorbed phage are transferred from wells labelled "A" to wells labelled "B". These are allowed to bind for 4 hrs at room temperature. The unbound phage are carefully pipetted away and each well is washed 4 times with 300  $\mu$ l wash buffer. Care is taken not to cross contaminate the wells.

Elution of DNA binding phage is done by the addition of 60  $\mu$ l 50 mM glycine-HCl pH 2.0 buffer for 10 minutes at 65°C. The eluted phage are transferred to a 60  $\mu$ l solution of 200 mM sodium phosphate, pH 7.4 to neutralize. One round of panning is now complete.

At this point it is preferred to amplify the phage isolated from DNA coated wells. This population of phage is not a pure population of phage that bind specific DNA target sequences, but rather a population that is substantially enriched for phage that bind to the target sequences. In the present example, four sublibraries are being isolated, one that binds to a NFIL6 site and three that have specificity toward different NF- $\kappa$ B sites. See Section 6.1.1 for a description of the oligonucleotide target ligands for these sublibraries. These sub-libraries are composed of "DNA binding phage" or DB phage.

Each sublibrary of DB phage is amplified, titered and stored, e.g., as described in Kay et al., 1993, Gene 128:59-65. To obtain phage that have the highest affinity for each of the DNA binding targets, three more rounds of panning are carried out in

substantially the same fashion as described above. However, since the population of DNA binding phage have been significantly enriched there are many copies of each unique phage in each  $10^{11}$  aliquot of the amplified sublibrary. Three rounds of additional panning toward each target is carried out without additional amplification of the phage in between adsorption and elution. After the final elution and neutralization, the phage are used to infect *E. coli* strain DH5 $\alpha$ F' and plated out to obtain substantially pure plaques.

About 24 individual plaques panned against each target sequence are picked and used to grow up a stock of 2 ml of phage. The TSAR insert in each phage is sequenced by the usual methods, e.g., Sanger dideoxy sequencing. Generally one obtains several identical or very similar phage in each group of 24 from one panning experiment. The translation of the DNA sequence (syngene) encoding the TSAR domain within the phage is the peptide sequence responsible for the binding of a given phage to the DNA target sequence. One isolate for each unique peptide encoding sequence is utilized for further characterization of its DNA binding properties.

25

#### 6.1.3. CHARACTERIZATION OF BINDING PHAGE

The selectivity of a given phage for a target sequence can be determined by any of several means. One method involves the carrying out of a "phage" ELISA utilizing an ELISA kit from Pharmacia (Product #27-9402-01) composed in part of horseradish peroxidase conjugated sheep anti-M13 antibody. The phage ELISA is carried out as follows:

1. The phage to be tested is grown in appropriate cells for 6 hrs to overnight in 2XYT at 37°C;

35

2. The cells are spun down and the supernatant is collected. This supernatant can be stored at 4°C for several days;
3. The wells of microtiter plates are  
5 coated with the oligonucleotides the phage is to be tested against. Coating can be done as in the panning methods described above;
4. About 100 µl of blocking solution (1 mg/ml BSA in 100 mM NaHCO<sub>3</sub>, pH 8.3) is added and the  
10 wells are incubated for 1 hr;
5. The liquid is flicked out and the wells are washed in PBS with 0.1 % Tween ® 20;
6. 50 µl of phage supernatant is added to the wells;
- 15 7. The wells are incubated at room temperature for 2 hrs;
8. The wells are washed multiple times with PBS with 0.1 % Tween ® 20. One wash should be for about 10 minutes;
- 20 9. 50 µl of Pharmacia's horseradish peroxidase conjugated sheep anti-M13 IgG (diluted 1:3000 in PBS with 0.1 % Tween ® 20 and 1 mg/ml BSA) is added and incubated for 1 hr at room temperature;
10. The wells are washed multiple times  
25 with PBS with 0.1 % Tween ® 20;
11. 100 µl of ABTS® (2,2'-Azino-di-[3-ethylbenzthiazdine sulfonate (6)] (0.2 mg/ml) is added;
12. Color develops in 10 to 30 minutes and  
30 absorbance at 405 nm is read in a microtiter plate reader.

Thus, by following the instructions provided by the manufacturer and utilizing ABTS® as substrate for the peroxidase, one can detect the amount of phage  
35 bound to a well in a multi-well plate in a quantitative manner. The binding selectivity for a given DNA target sequence of any unique TSAR phage can

be rapidly determined by bringing aliquots ( $10^8$  phage particles) of a particular phage into contact with specific target DNAs immobilized in wells of a multi-well plate. By using each of the 4 pairs of DNA targets described above in Section 6.1.1, appropriately immobilized, one can rapidly determine which phage binds specifically to which DNA sequence. A phage that binds to only one target sequence is a highly specific DNA binding phage (HSDB phage). A phage that binds to more than one target sequence is a cross-specific DNA binding phage.

Further characterization of a given peptide displaying phage's ability to bind to a particular target sequence can be determined by carrying out so-called competition experiments routinely used by those skilled in immunoassays. In these assays an aliquot of phage is first brought into contact with diverse pairs of non-biotinylated oligonucleotides under conditions and for a time sufficient to allow any binding to occur. Subsequently each aliquot is added to a well of a multi-well plate coated with a specific DNA target sequence. If the unlabeled competitor DNA is unable to bind to the TSAR domain of the phage, then the phage may bind to the specific DNA target sequence immobilized in the well. The phage bound in the well can be detected and measured by any convenient technique, e.g., use of the horseradish peroxidase conjugated sheep anti-M13 IgG and ABTS® of the phage ELISA described above. If the unlabeled DNA fragment is able to bind to the TSAR domain in the phage, then that phage will be unable to bind to the target DNA immobilized in the well. By varying the concentration of unlabeled DNA in solution one can estimate the relative affinity of a given phage TSAR domain for a specific DNA target as well as make a determination about the actual specificity of a TSAR domain for a specific DNA site.

It will be apparent to one skilled in the art that other means of analyzing the HSDB phage exist. It should be noted that the use of the binding competition experiments outlined earlier is very good for doing this as one can get good estimates of binding coefficients as well as dissociation constants of the HSDBs for targets and/or unrelated DNA sequences.

10        6.1.4. PHAGE CONTAINING BINDING DOMAINS SPECIFIC FOR NF- $\kappa$ B AND NF-IL6 BINDING SITES

HSDB phage that recognized the oligonucleotide formed by the sequences H2 $\kappa$ B-U and H2 $\kappa$ B-L (H2 $\kappa$ B oligonucleotide, see Section 6.1.1 above) were isolated from an R26 peptide display library. (See Section 6.9.4 and Figure 9 regarding the R26 library).

The R26 library was screened and panned as described above using "Universal Covalent" multiwell plates with the H2 $\kappa$ B oligonucleotide. Four phage clones were isolated that exhibited specific binding to the H2 $\kappa$ B oligonucleotide. To examine further the specificity of binding of these clones, they were tested in a phage ELISA (as described above) against the H2 $\kappa$ B oligonucleotide and three other oligonucleotides. The other oligonucleotides that were tested against these four phage were: (1) the oligonucleotide formed by IL6 $\kappa$ B-U and IL6 $\kappa$ B-L (IL6 $\kappa$ B oligonucleotide); (2) the oligonucleotide formed by IL8 $\kappa$ B-U and IL8 $\kappa$ B-L (IL8 $\kappa$ B oligonucleotide); (3) the oligonucleotide formed by NFIL6-U and NFIL6-L (NFIL6 oligonucleotide, see Section 6.1.1).

Figure 4 shows the results of these ELISAs. Clones 1, 2, and 6 all showed strong binding to the H2 $\kappa$ B oligonucleotide as compared to their binding to background (the BSA coated plates). Clone 5 showed less strong, but still about 2.5-fold greater, binding



to the H2κB oligonucleotide than to background. This binding on the part of clones 1, 2, 5, and 6 can be contrasted with the binding of m663 and a randomly selected phage from the R26 library, both of which  
5 bound the BSA coated plates about as well as they bound the plates coated with the H2κB oligonucleotide.

The exquisite specificity of binding of HSDBs is also evident in Figure 4. Clones 1, 2, 5, and 6 all bound the H2κB oligonucleotide far better  
10 than they bound any of the other three oligonucleotides. This is seen by the much higher ratios of binding to the oligonucleotide coated plates versus binding to the BSA coated plates when the oligonucleotide was the H2κB oligonucleotide rather  
15 than any of the other three oligonucleotides.

The oligonucleotide sequences of the random inserts of clones 1, 2, 5, and 6 were determined. This allowed for the determination of the corresponding amino acid sequences for the peptides  
20 encoded by these inserts. The following were the results obtained:

Clone 1 - SWCTYSGYCRVSSAGTAQRTSVDRDGM (SEQ  
ID NO: 143)  
25 Clone 2 - RTGNEQPPGSFGRAAGCFHPGCKYMKLN (SEQ  
ID NO: 144)  
Clone 5 - SDKYFHDIRKYHPAAATSYKTRPDMPST (SEQ  
ID NO: 145)  
Clone 6 - RTGNEQPPGSFGRAAGCFHPGCKYMKLN (SEQ  
30 ID NO: 144)

Note that the sequences of clones 2 and 6 are identical. In the discussion that follows, clone 1 is called H2κB-1 and clone 2 is called H2κB-2.

35 HSDB phage that recognized the oligonucleotide formed by the sequences NFIL6-U and NFIL6-L (NFIL6 oligonucleotide, see Section 6.1.1

above) were isolated from the DC43 peptide display library. (See Section 6.9.6 and Figure 13 regarding the DC43 library).

5 The DC43 library was screened and panned as described above using "Universal Covalent" multiwell plates with the NFIL6 oligonucleotide. A phage clone (NFIL6-1) was isolated that exhibited specific binding to the NFIL6 oligonucleotide.

10 The oligonucleotide sequence of the random insert of NFIL6-1 was determined. This allowed for the determination of the corresponding amino acid sequence for the peptide encoded by this insert. The following was the result obtained:

15 NFIL6-1 - REWGVPGAHNRI RDHCNGPRCHAIRTNASHTQHI  
SRPPD (SEQ ID NO: 146)

The selectivity of binding of the phage H2κB-1, H2κB-2, and NFIL6-1 toward the three  
20 oligonucleotides H2κB, IL6κB, and NFIL6 was tested by phage ELISA. Phage ELISA were performed as in Section 6.1.3. Each phage was assayed for its ability to bind to each oligonucleotide following immobilization of that oligonucleotide in the well of a microtiter  
25 plate. This binding was compared to the binding of the phage to wells that had been coated with BSA (bovine serum albumin). As a negative control, the binding of the parental phage m663 to each of the oligonucleotides was also assayed. The results are  
30 shown in Figure 14.

Figure 14 shows that libraries of random peptides can be successfully screened to identify and isolate binding domains that specifically bind to specific DNA sequences. Phage NFIL6-1 binds well to  
35 the NFIL6 oligonucleotide with some binding to the other two oligonucleotides. Phage H2κB-2 binds only to the target H2κB oligonucleotide. Phage H2κB-1

shows much greater binding to the H2κB oligonucleotide than to the other oligonucleotides.

The target specificity of H2κB-2 was investigated further by testing its ability to bind to DNA sequences that are variants of the H2κB oligonucleotide. New DNA oligonucleotides were synthesized; the sequences of these oligonucleotides deviated from the sequence of the H2κB oligonucleotide by only one or two (and in one instance, three) bases. Oligonucleotides for the upper and lower strands were synthesized. For each pair of upper and lower strands, 20 μg of upper and lower oligonucleotides were combined in 100 μl of TE buffer supplemented with 200 mM NaCl. Annealing and filling-in was carried out at 65°C for 5 minutes, 42°C for 5 minutes, and 37°C for 15 minutes. The resulting double stranded DNA fragments were diluted in PBS, pH 5.0. 50 μl of the appropriate DNA fragments were added to Costar Universal Covalent microtiter plates as in Section 6.1.2 and incubated overnight at 4°C. Phage ELISA assays were carried out three times for each target DNA using the H2κB-2 phage. The results are shown in Table 8. For each variant DNA target, the binding of H2κB-2 to that target was compared to the binding of H2κB-2 to the H2κB oligonucleotide (wild type (WT)) and expressed as percent binding compared to the wild type oligonucleotide (% WT).

TABLE 8  
Specificity of H2k $\kappa$ B-2 Binding to DNA Targets

	Mutant	H-2 Kappa B Regulatory Sequence																		% WT	
5	wt	G	G	G	T	G	G	G	G	A	T	T	C	C	C	C	A	T	C	T	100
	M-1	T	T																		78
	M-2			T	G																129
10	M-3					T	T														139
	M-4							T	T												17
	M-4a							T													32
	M-4b								T												132
15	M-5									C	G										156
	M-5a									C											137
	M-5b										G										46
	M-4/5+								T	C											173
	M-4/5-							T			G										32
20	M-6											G	A								51
	M-6a											G									135
	M-6b												A								130
	M-7													A	A						101
25	M-8															A	C				99
	M-9																	G	A	G	84
	Consensus oligo	A	A	A	T	G	T	G	G	A	T	T	C	C	C	T	C	T	G	G	13

30

35

Inspection of the data of Table 8 reveals that a core group of "critical" residues (those whose mutation results in decreased binding) contributes significantly to the H2k $\kappa$ B-2 phage's ability to bind DNA. For example, the G at position 7, when mutated to T (with no other changes), results in diminished

binding (32% of wild type). It is clear, however, that the context of the critical residues is also very important. When the G to T mutation at position 7 is coupled to a G to T mutation at position 8 (see mutant M-4), binding is further decreased to 17% of wild type. Yet when the G to T mutation at position 8 occurs alone, binding is increased to 132% of wild type (see mutant M-4b). A "consensus" DNA fragment which has all the apparent critical residues, but is flanked with a random array of nucleotides, is virtually unable to bind the H2xB-2 phage clone.

#### 6.1.5. PRODUCTION AND ANALYSIS OF MUTAGENIZED BINDING PHAGE

For the present invention, the relative specificity of an HSDB for a specific DNA site is more important than the actual affinity of the domain for the DNA site. In gene therapy, the effect of low affinity can be compensated for by increased levels of gene expression within the target cell. However, it is very important that the specificity be high in order to bring about the desired activation or inhibition of the correct gene and not others.

If, after analysis of the DNA binding phages isolated by the protocol outlined above, it is desirable to have HSDBs with greater selectivity, then the amino acid sequence encoded by the TSAR domain of the phage having the best properties can be mutagenized by conventional means known to those skilled in the art. In particular, saturation mutagenesis is carried out using synthetic oligonucleotides synthesized from "doped" nucleotide reservoirs. The doping is carried out such that the original peptide sequence is represented only once in  $10^6$  unique clones of the mutagenized oligonucleotide. The assembled oligonucleotides are cloned into the parental TSAR vector. Preferably the vector is m663

(Fowlkes et al., 1992, BioTechniques 13:422-427). m663 is able to make blue plaques when grown in *E. coli* strain JM101 or DH5 $\alpha$ F'. A library of greater than  $10^6$  is preferred; however a library of  $10^5$  is  
5 sufficient to isolate TSAR phage displaying peptide domains with increased selectivity for binding to the target DNA sequence.

Once a DNA binding TSAR mutant library has been constructed and amplified, it can be panned with  
10 immobilized target DNA sequences in a manner analogous to that described for the isolation of the initial DNA binding phage. It is preferable, but not necessary, to selectively pan out those phage capable of binding to sequences related to the target DNA sequence.  
15 Preferably, the related sequence is that of another site subject to regulation by the same transcription factor as the site desired to be regulated by the syngene product. Therefore, it is preferable to use four wells for this set of panning and phage  
20 isolations. Well "A" is coated only with streptavidin; well "B" is coated with a DNA binding target that is totally unrelated to the relevant DNA target; well "C" is coated with a DNA binding target that is related to the relevant DNA target; and well  
25 "D" is coated with the relevant DNA target. For the actual panning, an aliquot of the mutant TSAR DNA binding phage ( $50\ \mu\text{l}$ ,  $10^{11}$  phage/ml) is added first to well "A", incubated for 2 hrs, then transferred to well "B", incubated for 2 hrs, then transferred to  
30 well "C" for 2 hrs and finally, transferred to well "D" for 2 hrs. Well "D" is washed 4 times with wash buffer and then eluted as described above. The neutralized phage are used to infect *E. coli* and stocks of the more specific DNA target binding phage  
35 are prepared. From these stocks one repeats another 3 rounds of panning without intervening phage amplification steps. The phage obtained by this

process are called "highly specific DNA binding phage" or HSDB phage.

Another means for blocking the binding of phage capable of binding to sequences related, but not identical, to the target DNA sequence is to add soluble DNA oligonucleotide fragments (to a final concentration of 0.1  $\mu$ g/ml) to aliquots of the mutant library before panning. The added DNA oligonucleotide fragments bear the sequences related (but not identical) to the target sequence.

To characterize HSDB phage relative to parental and other DB phage isolated from a TSAR-9 library, one takes advantage of the fact that the TSAR-9 library phage have no lacZ activity, whereas the mutant libraries used to clone the HSDB phage are from a vector that expresses lacZ activity in the appropriate *E. coli* cell lines. To determine if a mutant binds to the target DNA with more specificity than its parent DB phage, equal amounts of the two phage to be compared are mixed together and then applied to a well containing a target DNA sequence. In addition, an equal amount of the two phages are applied to a well containing another DNA sequence. After the appropriate washing and elution, the phage recovered are plaqued onto the appropriate *E. coli* in the presence of X-gal. Ratios of blue to white phage are calculated. Since the mutant phage can convert X-gal to a blue product, one can readily determine which mutant phage have improved DNA binding specificity relative to DB phage isolated from the TSAR-9 library. The mutants that bind less well to the irrelevant DNA target are selected for characterization for potential as therapeutic syngene targeting domains.

To determine critical residues for the binding of H2K-B-2 phage to its DNA target and to determine if the affinity or specificity of binding of

this phage could be enhanced, a scheme was developed to sample a large number of phage variants based on the peptide encoded by H2κB-2.

- Oligonucleotides o8909 and o8545 were
- 5 synthesized and purified by reverse phase HPLC (see Figure 15). Double stranded DNA fragments were made by annealing 10 μg of each oligonucleotide in 200 μl of 1X Sequenase<sup>™</sup> buffer (United States Biochemical Corp., Cleveland, OH) with 10 mM DTT and 150 μM of
- 10 each dNTP. The fill-in reaction was initiated with 3 μl of Sequenase<sup>™</sup>, Version 2.0 (United States Biochemical Corp., Cleveland, OH) and carried out at 37°C for 30 minutes. The filled-in, double stranded fragments were phenol/chloroform extracted and ethanol
- 15 precipitated. The dried pellet was resuspended in 80 μl of TE and digested with Xba I and Xho I (New England Biolabs, Beverly, MA) according to the supplier's recommendations. The appropriate sized DNA fragments were purified by polyacrylamide gel
- 20 electrophoresis (PAGE). The fragments were cloned into m663 in the usual manner. A library of about 1.5 x 10<sup>8</sup> phage variants was obtained. This library is called the ME#1 library (See Figure 15). The ME#1 library was panned using the original H2κB
- 25 oligonucleotide target as well as with the IL6κB, IL8κB, and NFIL6 oligonucleotides. A large number of phage were identified that bound specifically to the H2κB oligonucleotide; no phage were identified that bound to any of the other targets.
- 30 Twenty-two clones from the ME#1 library that bound the H2κB oligonucleotide (binders) were selected for further analysis. Also, twenty-six clones were identified that did not have significant ability to bind the H2κB oligonucleotide (non-binders). In
- 35 addition, random plaques from the ME#1 library were tested by phage ELISA for the ability to bind the H2κB



oligonucleotide; about one third demonstrated some binding ability.

The amino acid sequences of the inserts of the binders and non-binders from the ME#1 library were determined. The sequences were analyzed to determine the frequencies at which each amino acid appeared at each position for the binders and for the non-binders. The results are shown in Table 9.

TABLE 9  
Molecular Evolution of H2kB-2:  
Frequency of Residues for Binding and Non-Binding Members

Residue	Binding Sequences						Non-binding Sequences				
	Original Residue			Other Residues			Original Residue		Other Residues		
	p	Actual		p	Actual		p	Actual	Other	p	%
1. R	94%	61%	S	8%	36%		94%	81%	S	8%	19%
2. T	77%	86%					77%	77%			
3. G	77%	95%					77%	69%			
4. N	51%	55%					51%	65%			
5. E	25%	45%	D	51%	45%		25%	32%	D	51%	69%
6. Q	25%	23%	H	51%	59%		25%	38%	H	51%	35%
7. P	77%	82%					77%	81%			
8. P	77%	82%					77%	57%			
9. G	77%	82%					77%	81%			
10. S	51%	91%					51%	73%			
11. F	51%	45%	L	19%	23%		51%	46%	L	22%	35%
12. G	77%	73%					77%	85%			
13. R	75%	73%					75%	65%			
14. A	77%	77%					77%	58%			
15. A	25%	14%	S	25%	41%		25%	27%			
16. G	100%	100%					100%	100%			
17. C	17%	100%					17%	31%			
18. F	51%	86%					51%	35%			
19. H	51%	100%					51%	27%			
20. P	77%	91%					77%	85%			
21. G	77%	95%					77%	77%			
22. C	51%	100%					51%	31%			
23. K	24%	95%	N	51%	5%		24%	12%	N	51%	58%
24. Y	25%	100%	S	3.5%	0%		25%	27%	S	3.5%	31%
25. M	25%	18%	I	51%	82%		25%	15%	I	51%	73%

	K	25%	73%	N	51%	5%	25%	12%	N	51%	35%
26.				H	2.4%	18%			H	2.4%	0%
27.	L	25%	18%	H	2.4%	73%	25%	62%	H	2.4%	0%
28.	N	51%	18%	K	25%	73%	51%	50%	K	25%	12%

5

In Table 9, "Residue" indicates the residues in that position in the insert of the original H2xB-2 clone from the amino terminal to the carboxy terminal position; "p" indicates the percentage of clones expected to have the original residue occurring at a given position based on the scheme utilized for synthesizing the oligonucleotides for the ME#1 library; "Actual" indicates the frequency that a specific residue was observed to occur in a specific position.

Table 9 shows that there was no apparent bias in the sequences for both the binders and non-binders in the first 16 residues. However, at position 17, 100% of the binders have the original cysteine, but only 31% of the non-binders do. For both groups, this is more than the expected frequency of 17%. It appears that this cysteine is critical for the binders, but that it is also selected for in the library as a whole.

There are significant differences between the two groups for residues 18 through 28. While the original residue is clearly favored in the binding class for residues 18-24 and 26, this is not the case for positions 25, 27, and 28. In position 25, the original methionine was observed at less than the expected frequency in both the binders and non-binders and isoleucine was observed at higher than expected frequency in both classes.

Residues 27 and 28 are very informative. For the binders, histidine was observed at position 27

in 73% of the clones, but was never observed at that position in the non-binders. At position 28, lysine was observed in 73% of the binders but in only 12% of the non-binders. Furthermore, the original asparagine was expected at a frequency of 51%, but occurred in only 18% of the binders.

Thus, while the H2κB-2 clone clearly binds well to the H2κB oligonucleotide target, it does not carry the optimal amino acid sequence to do so (see below and Figures 16A and 16B for a further discussion of this point). Most important, it was learned that substituting a histidine for the lysine at position 26 and a lysine for the asparagine at position 28 leads to avid binding of the clones to the target DNA.

Phage stocks were prepared for the binders from the ME#1 library as well as for the original H2κB-2 clone and the stocks were titered by serial dilution. Subsequently, the appropriate dilutions of each phage were analyzed by phage ELISA for binding to the H2κB oligonucleotide. The results are shown in Figures 16A and 16B. Binding is expressed as signal strength (O.D.). Phage having higher relative avidity for the target (as compared to H2κB-2) continue to produce a detectable signal at low concentrations of phage and thus the curves for those phage are shifted to the left as compared to the curve for H2κB-2.

It is readily evident from Figures 16A and 16B that most binders from the ME#1 library have a greater avidity for the target H2κB oligonucleotide than does the original H2κB-2 clone. Of the 22 binders analyzed, 5 are less avid binders than H2κB-2 and many have comparable binding avidity; however, over one third show avidity of more than two orders of magnitude greater than that of H2κB-2. Table 10 shows the amino acid sequences of the binders tested and summarizes their binding avidity.

TABLE 10

## Binding Strength of Clones from the ME#1 Library

			SEQ ID
5	STRONG BINDERS		NO:
	950406-10 (C1-9)	SIDKVQPPGTSGRKTGCFHPGCKYIKLK	147
	950406-8	STGKEHPPGSFGRATGCFHPGCKYIKLK	148
	950406-11	STGKEHSPGSLGRAPGCWHPGCKYIKLK	149
	950406-1	RTGNDQPPRSFGHATGCFHPGCKYIKNK	150
	950406-2	STGKDHPSSFGRAAGCFHPGCKYIKHK	151
10	C1-8	RTGEDHPPGSFSGKAAGCFHPGCKYIKHK	152
	D1-11	RTGNDHPSVSYGRASGCFHAGCKYIKHK	153
	D1-1	RTGNENPPGSLGRGTGCFHAACKYIKHK	154
	F15-D1	RAGTDQPPGSFVRASGCFHPGCKYIKHK	155
	MEDIUM BINDERS		
	F15-A9b	RTGVDQSPESFDRATGCFHPGCKYIKHN	156
15	F15-B2b	RTGNDDPTGALARNSGCYHPGCKYINHK	157
	F15-A4	RTGNDNPPGSSGPASGCFHPGCKYMKHK	158
	C1-11	STGKEHPPGSFGRVSGCFHPGCKYMKHN	159
	C1-6	RTGNDHPSGSFEHAAGCFHPGCKYMKHK	160
	D1-3	RTGDDHPTGSIGRASGCFHPGCKYIKHN	161
	H2kB-2	RTGNEQPPGSFSGRAAGCFHPGCKYMKLN	144
20	WEAKER BINDERS		
	B1-12	STGKEHPPGSIGRASGCFHPGCKYIRHT	162
	D1-7	RTGNEHPPGSIGRASGCFHPGCNYIKHK	163
	F15-B3	RTGNEHPPGSLGRAPGCYHPGCKYIHHP	164
	F15-A7	RTGNYDPPGSFGPSTGCFHPGCKYIHHK	165
	A1-13	STGNEQHPSGLRRASGCFHPGCKYIHHN	166
25	F15-B7	STGNEHPPGSVGHAPGCFHPGCKYMHK	167

These results demonstrate that the above strategy for evolving a peptide sequence isolated from a random peptide library towards one with greater binding avidity for its target was highly successful.

A group of the binders from the ME#1 library was tested by phage ELISA for the ability to bind to the same variant H2kB oligonucleotide targets as were described in Table 8. H2kB-2 was similarly tested. In order to compare the binding of these clones, the O.D. for binding to the original H2kB oligonucleotide

target (WT) was normalized to 100%. Table 11 shows the results.

TABLE 11

Binding of Clones from Molecular Evolution Library #1 to H2kB Variant Targets

Clones	H2kB-2	950406-10	F15-D1	F15-B3	F15-A7	F15-B2
O.D.	0.527	1.021	0.815	0.117	0.257	0.480
wt	100	100	100	100	100	100
M-1	48	82	67	82	95	63
M-2	75	76	69	85	86	57
M-3	106	76	93	111	86	143
M-4	16	24	10	47	46	16
M-4a	19	30	17	55	54	12
M-4b	94	52?	59	64	71	103
M-5	184	104	146	366	258	198
M-5a	175	116	154	276	249	206
M-5b	53	66	60	67	73	47
M-4/5+	195	107	133	503	286	209
M-4/5-	25	57	33	61	73	27
M-6	74	79	79	71	44	56
M-6a	161	115	129	235	179	163
M-6b	155	103	119	194	169	182
M-7	102	76	92	77	79	70
M-8	126	83	105	122	150	162
M-9	94	89	108	80	105	75
conc.	11	2	4	46	37	8
BSA	6	3	5	29	9	6

All of the binders that had previously been shown to have increased avidity for the H2kB oligonucleotide (as compared to H2kB-2) maintained the same relative target specificity as H2kB-2. However, some binders with less avidity for the H2kB oligonucleotide had increased avidity for certain of the variant H2kB oligonucleotide targets.

Since the critical residues identified in the first round of molecular evolution were at the carboxy terminal end of the expressed peptide insert,

it was not possible to rule out some contribution from residues in the phage pIII protein to the DNA binding activity. A scheme was developed to construct additional molecular evolution libraries for panning with the H2xB oligonucleotide. This was done to ensure that the critical DNA binding residues would be able to function in the context of other gene sequences, e.g., nuclear localization sequences, transcriptional activation sequences, for constructing a functional syngene directed to the regulation of the activity of a gene with H2xB sequences.

Two libraries were constructed. Molecular evolution library #2a (the ME#2a library) contained the critical residues as a fixed core flanked by 10 random residues on each side. Molecular evolution library #2b (the ME#2b library) was constructed in a similar manner, but with a smaller group of core residues. Specifically, the ME#2b library lacks the initial arginine, serine, glycine, and arginine found within the core of the ME#2a library. In addition, the core of the ME#2b library was flanked with 12 random residues on each side. Figures 17A and 17B summarize the construction of the ME#2a and ME#2b libraries.

The ME#2a and ME#2b libraries were panned using the H2xB oligonucleotide as ligand. A number of clones were isolated as binders from the ME#2a and ME#2b libraries. Phage stocks were prepared for the binders from the ME#2a and ME#2b libraries as well as for the original H2xB-2 clone and the stocks were titered by serial dilution. Subsequently, the appropriate dilutions of each phage were analyzed by phage ELISA for binding to the H2xB oligonucleotide. The results are shown in Figure 18. Binding is expressed as signal strength (O.D.). Phage having higher relative avidity for the target (as compared to H2xB-2) continue to produce a detectable signal at low

concentrations of phage and thus the curves for those phage are shifted to the left as compared to the curve for H2kB-2.

While many clones from the ME#2a and ME#2b libraries did not have enhanced binding activity as compared to H2kB-2, it is apparent that a significant number have an apparent avidity two to three orders of magnitude greater than that of H2kB-2. However, none of the clones from the ME#2a or ME#2b libraries had an avidity as great as that of clone 959496-10 from the ME#1 library.

Table 12 shows the amino acid sequences of the inserts of the binders from the ME#2a and ME#2b libraries as well as the inserts of clone 959496-10 and H2kB-2.

TABLE 12

		SEQ ID NO.
	J5-1A7	SADLVAPSGENRSGRGCFHPGCKYIKHKPTTPPAPSSA
20	J5-2A7	SAATADLGVGERSGRGCFHPGCKYIKHKGSTSQSQPL
	J5-1A10	SPDTKSPGGDVRSGRGCFHPGCKYIKHKSDSSHKEATT
	J5-2A4	SVTSDDPQGRERSGRGCFHPGCKYIKHKQHATSESPNL
	J5-2A8	SGREPEGPSEQRSGRGCFHPGCKYIKHKTQSTMHTTPA
	J5-2B10	SFRALDEPTATRSRGCFHPGCKYIKHKSPKDSFPNTTAA
	J5-1B2	SGTVPTAEVNNRPGCFHPGCKYIKHKLPRQPSPTPPLA
	J5-1B6	SDWSERYGGERGGCFHPGCKYIKHKNKNSGTPLPSDP
	J5-1B1	SNTDTAWAGNGDRGCFHPGCKYIKHKTPTPTNKTYETP
25	J5-2B3	SGNVDNTERGGKGCFFHPGCKYIKHKRPNAYEVVPPLD
	J5-2B2	SERAERGEWNWSAGCFHPGCKYIKHKSPRGANRSLVGA
	J5-2B4	SSDGDTPGGSRKGGCFHPGCKYIKHKPSQLPREGTHN
	950406-10	SIDKVQPPGTSGRKTGCFHPGCKYIKLK
	H2kB-2	RTGNEQPPGSFGRAAGCFHPGCKYMKLN

It can be seen from Table 12 that certain residues are highly favored in the sequences flanking the critical core residues. This is particularly apparent in the clones from the ME#2b library which lacked the initial arginine, serine, glycine, and arginine residues found in the core of the ME#2a library. In all clones from the ME#2b library that

bound well to the target DNA there are charged residues in the four residues upstream of the core.

6.2. ISOLATION OF HSDBs FOR GATA  
TRANSCRIPTION FACTOR BINDING SITES

5

The methods described in Section 6.1 and its subsections above for the isolation of HSDBs for NF- $\kappa$ B and NF-IL6 binding sites can be easily modified to isolate HSDBs for GATA transcription factor binding sites.

10

The major modification consists in the use of DNA target sequences from GATA transcription factor binding sites rather than DNA target sequences from NF- $\kappa$ B or NF-IL6 binding sites. For example, for the GATA-2 site in the human preproendothelin-1 gene, the target DNA sequence is 5' ctggccTTATCTccggct (SEQ ID NO: 180) for the upper strand and 5'agccggAGATAAggccag (SEQ ID NO: 181) for the lower strand (Dorfman et al., 1992, J. Biol. Chem. 267:1279-1285).

15

20

The human gastric (H<sup>+</sup> + K<sup>+</sup>) ATPase gene is responsible for maintaining the large (about 5 units) pH difference between the cytoplasm and gastric lumen in the stomach. It has been proposed that regulation of the human gastric (H<sup>+</sup> + K<sup>+</sup>) ATPase gene would be useful in the treatment of gastric ulcers (Maeda, 1994, J. Biochem. 115:6-14). It would be useful to have syngenes that encode HSDBs that could be used to modulate the activity of the human gastric (H<sup>+</sup> + K<sup>+</sup>) ATPase gene. Such HSDBs can be obtained by screening random peptide libraries with the an oligonucleotide formed from an upper strand of 5'gacatGGGGGATCTGGgca (SEQ ID NO: 182) and a lower strand of 5'tgcCCAGATCCCCCcatgtc (SEQ ID NO: 183). This oligonucleotide represents a sequence similar to that of bindings site for the GATA-GT family of transcription factors (Maeda et al., 1990, J. Biol. Chem. 265:9027-9032).

25

30

35



For the GATA-3 site in the human T cell receptor  $\delta$  gene enhancer, the target DNA sequence is 5'gacactTGATAAcagaaa (SEQ ID NO: 184) for the upper strand and 5'tttctgTTATCAagtgtc (SEQ ID NO: 185) for the lower strand (Ko et al., 1991, Mol. Cell. Biol. 11:2778-2784).

As for NF- $\kappa$ B and NF-IL6 transcription factor binding site HSDBs, the DNA target sequences for the isolation of GATA transcription factor binding site HSDBs can be monovalent or multivalent. Panning, amplification, and analysis of the isolated phage are done as for NF- $\kappa$ B or NF-IL6. Mutagenesis to produce HSDBs of greater selectivity can also be done as for NF- $\kappa$ B or NF-IL6.

### 6.3. ISOLATION OF HSDBs FOR AP-1 TRANSCRIPTION FACTOR BINDING SITES

The transcription factor AP-1, which consists of a heterodimer of the products of the c-fos and c-jun proto-oncogenes, is involved in the transcriptional regulation of a number of genes. For example, an AP-1 binding site is a component of the TPA (12-O-tetradecanoyl-phorbol-13-acetate)-responsive enhancer element that is involved in conferring serum-responsive transcriptional activity to many genes (Bohmann et al., 1987, Science 238:1386-1392; Angel et al., 1988, Nature 332:166-171). It would be of great value to have syngenes that encoded products that could specifically bind to AP-1 sites, since this would allow for the regulation of many genes that are involved in the growth response of cells.

The methods described in Section 6.1 and its subsections above for the isolation of HSDBs for NF- $\kappa$ B and NF-IL6 binding sites can be easily modified to isolate HSDBs for AP-1 transcription factor binding sites.

The major modification consists in the use of DNA target sequences from AP-1 transcription factor binding sites rather than DNA target sequences from NF- $\kappa$ B or NF-IL6 binding sites. In general, the target oligonucleotides will have the following sequences:  
5'xxxxxTGA(G/C)T(C/A)Axxxxx (SEQ ID NO: 186) for the upper strand and 5'xxxxxT(G/T)AC/G)TCAxxxxx (SEQ ID NO: 187) for the lower strand. See Gambari and Nastruzzi, 1994, Biochemical Pharmacology 4:599-610.  
Where two nucleotides are shown in parenthesis, the choice of which nucleotide to use will depend upon which specific AP-1 binding site it is desired to isolate an HSDB for. Similarly, the choice of which nucleotide to use in the positions marked by an x will depend upon the nucleotide at those positions in the specific AP-1 binding site it is desired to isolate an HSDB for.

As for NF- $\kappa$ B and NF-IL6 transcription factor binding site HSDBs, the DNA target sequences for the isolation of AP-1 transcription factor binding site HSDBs can be monovalent or multivalent. Panning, amplification, and analysis of the isolated phage are done as shown above for NF- $\kappa$ B or NF-IL6. Mutagenesis to produce HSDBs of greater selectivity can also be done as for NF- $\kappa$ B or NF-IL6.

#### 6.4. ISOLATION OF HSDBs FOR ATF TRANSCRIPTION FACTOR BINDING SITES

The methods described in Section 6.1 and its subsections above for the isolation of HSDBs for NF- $\kappa$ B and NF-IL6 binding sites can be easily modified to isolate HSDBs for ATF transcription factor binding sites.

The major modification consists in the use of DNA target sequences from ATF transcription factor binding sites rather than DNA target sequences from NF- $\kappa$ B or NF-IL6 binding sites. In general, the target

oligonucleotides will have the following sequences:  
5'xxxxxTGACG(C/T)(C/A)(G/A)xxxxx (SEQ ID NO: 188) for  
the upper strand and 5'xxxxxT(G/T)AC/G)TCAXxxxxx (SEQ  
ID NO: 189) for the lower strand. See Gambari and  
5 Nastruzzi, 1994, Biochemical Pharmacology 4:599-610.  
Where two nucleotides are shown in parenthesis, the  
choice of which nucleotide to use will depend upon  
which specific ATF binding site it is desired to  
isolate an HSDB for. Similarly, the choice of which  
10 nucleotide to use in the positions marked by an x will  
depend upon the nucleotide at those positions in the  
specific ATF binding site it is desired to isolate an  
HSDB for.

As for NF- $\kappa$ B and NF-IL6 transcription factor  
15 binding site HSDBs, the DNA target sequences for the  
isolation of ATF transcription factor binding site  
HSDBs can be monovalent or multivalent. Panning,  
amplification, and analysis of the isolated phage are  
done as shown above for NF- $\kappa$ B or NF-IL6. Mutagenesis  
20 to produce HSDBs of greater selectivity can also be  
done as for NF- $\kappa$ B or NF-IL6.

#### 6.5. ISOLATION OF HSDBs FOR ETS TRANSCRIPTION FACTOR BINDING SITES

25 The methods described in Section 6.1 and its  
subsections above for the isolation of HSDBs for NF- $\kappa$ B  
and NF-IL6 binding sites can be easily modified to  
isolate HSDBs for binding sites for members of the Ets  
family of transcription factors.

30 The major modification consists in the use of DNA  
target sequences from specific Ets transcription  
factor binding sites rather than DNA target sequences  
from NF- $\kappa$ B or NF-IL6 binding sites. In general, the  
target oligonucleotides will have sequences determined  
35 by the specific nucleotide sequence of the specific  
Ets transcription factor binding site that it is  
desired to regulate.

For example, if it is desired to isolate an HSDB specific for the stromelysin gene, the target oligonucleotide will have an upper strand of 5'GCAGGAAGCA (SEQ ID NO: 190) and the lower strand will be 5'TGCTTCCTGC (SEQ ID NO: 191). This oligonucleotide corresponds to an Ets binding site in the stromelysin promoter (Waslyk et al., 1993, Eur. J. Biochem. 211:7-18).

If it is desired to isolate an HSDB specific for the T cell receptor  $\alpha$  gene, the target oligonucleotide will have an upper strand of 5'AGAGGATGTG (SEQ ID NO: 192) and the lower strand will be 5'CACATCCTCT (SEQ ID NO: 193). This oligonucleotide corresponds to an Ets binding site in the T cell receptor  $\alpha$  gene promoter (Waslyk et al., 1993, Eur. J. Biochem. 211:7-18).

If it is desired to isolate an HSDB specific for the c-fos gene, the target oligonucleotide will have an upper strand of 5'ACAGGATGTC (SEQ ID NO: 194) and the lower strand will be 5'GACATCCTGT (SEQ ID NO: 195). This oligonucleotide corresponds to an Ets binding site in the serum response element of the c-fos gene promoter (Waslyk et al., 1993, Eur. J. Biochem. 211:7-18).

As for NF- $\kappa$ B and NF-IL6 transcription factor binding site HSDBs, the DNA target sequences for the isolation of Ets transcription factor binding site HSDBs can be monovalent or multivalent. Panning, amplification, and analysis of the isolated phage are done as shown above for NF- $\kappa$ B or NF-IL6. Mutagenesis to produce HSDBs of greater selectivity can also be done as for NF- $\kappa$ B or NF-IL6.

6.6. ISOLATION OF BINDING DOMAINS  
SPECIFIC FOR LUMENAL PROTEINS  
OF THE ENDOPLASTIC RETICULUM

To isolate binding domains specific for lumenal proteins of the endoplasmic reticulum, use is made of the conserved tetrapeptide KDEL (SEQ ID NO: 116), which is found at the carboxy terminus of such proteins. Libraries such as, for example, the TSAR libraries described herein, are screened with a synthetic peptide ligand of the following sequence: KDELXXXXXX (SEQ ID NO: 196), where the identity of the positions marked with an X are determined by the specific amino acids at those positions in the specific lumenal proteins for which it is desired to isolate a binding domain.

The synthetic peptides can be synthesized by any of several well known methods in the art. Panning, amplification, and analysis of the isolated phage are done in a manner that is broadly similar to the manner in Sections 6.1.2 and 6.1.3, with the difference that the ligand used is a peptide rather than an oligonucleotide. Methods of screening are disclosed in Sections 5.2 and 6.1.2. Alternatively, any well known method of screening a peptide library using a peptide as a ligand may be used.

6.7. ISOLATION OF BINDING DOMAINS  
SPECIFIC FOR INTEGRAL MEMBRANE  
PROTEINS OF THE TRANS GOLGI NETWORK

To isolate binding domains specific for integral membrane proteins of the trans Golgi network, use is made of the tetrapeptide YQRL (SEQ ID NO: 117), which has been found to be necessary and sufficient to for the targeting of membrane proteins to the trans Golgi network (Bos et al., 1993, EMBO J. 12:2219-2228; Humphrey et al., 1993, J. Cell Biol. 120:1123-1135). Libraries such as, for example, the TSAR libraries described herein, are screened with a synthetic

peptide ligand of the following sequence: XXXYQRLXXX  
(SEQ ID NO: 197), where the identity of the positions  
marked with an X are determined by the specific amino  
acids at those positions in the specific proteins for  
5 which it is desired to isolate a binding domain.

The synthetic peptides can be synthesized by  
any of several well known methods in the art.  
Panning, amplification, and analysis of the isolated  
phage are done in a manner that is broadly similar to  
10 the manner in Sections 6.1.2 and 6.1.3, with the  
difference that the ligand used is a peptide rather  
than an oligonucleotide. Methods of screening are  
disclosed in Sections 5.2 and 6.1.2. Alternatively,  
any well known method of screening a peptide library  
15 using a peptide as a ligand may be used.

#### 6.8. BIOLOGICAL ACTIVITY OF SYNGENE-ENCODED PRODUCTS THAT BIND TO THE NF-IL6 AND NF- $\kappa$ B SITES

---

In order to demonstrate the biological  
20 activity and utility of the HSDB-containing syngenes,  
one can exploit the observation of Stein and Baldwin,  
1993, Mol. Cell. Bio. 13:7191-7198 that there is a  
functional and physical association of the NF- $\kappa$ B  
family members with NF-IL6 (also known as C/EBP $\beta$ )  
25 family members. The interaction of NF- $\kappa$ B with NF-IL6  
is such that, when both transcription factors are  
expressed, promoters with only NF- $\kappa$ B binding sites are  
inhibited but promoters with both NF- $\kappa$ B and NF-IL6  
binding sites are synergistically stimulated. NF- $\kappa$ B  
30 and NF-IL6 are activated by important inflammatory  
cytokines such as interleukin-1 and IL-6. Therefore,  
the interactions between NF- $\kappa$ B and NF-IL6 are likely  
to be involved in T-cell activation and the  
acute-phase response. Several genes having promoters  
35 that have closely spaced NF- $\kappa$ B and NF-IL6 binding

sites have been cloned. Examples are the IL-6 and the IL-8 genes.

Genes such as IL-6 and IL-8 that have both NF-IL6 and NF- $\kappa$ B sites are transcriptionally activated by the induction and expression of those transcription factors together. In addition, genes such as the MHC class I genes (for example H-2Kb) that have only an NF- $\kappa$ B site (and no NF-IL6 site) are transcriptionally antagonized by the expression of NF-IL6 and NF- $\kappa$ B together. They are, however, transcriptionally activated by the induction and expression of NF- $\kappa$ B alone. These observations allow one to readily determine the action of the syngenes that bind to the NF-IL6 and NF- $\kappa$ B sites when the appropriate syngenes are expressed in tissue culture cells along with appropriate reporter gene constructs.

The basic assay involves the introduction of reporter constructs, syngene constructs and plasmids expressing one or more transcriptional factors.

Reporter constructs: 1) Three copies of the H2 $\kappa$ B binding site fused upstream of a minimal fos promoter linked to the chloramphenicol acetyltransferase (CAT) gene (called pMHC-CAT herein, see Scheinman et al., 1993, Mol. Cell. Biol. 13:6089-6101)); 2) Plasmid pNFIL-6-CAT, which has a single copy of an oligonucleotide with a NFIL-6 site derived from the c-fos serum response element (5'-agcttgATTAGGACATcg3' (SEQ ID NO: 198) (binding site printed in upper case) in a HindIII-BamHI-cut pTATA-CAT (referred to as C/EBPbeta-TATA-chloramphenicol acetyltransferase reporter plasmid by Stein et al., 1993, Mol. Cell. Biol. 13:3964-3974); 3) Plasmid pIL-8-CAT, a IL-8 wild-type reporter construct generated by cloning a single copy of an oligonucleotide encompassing the bp -97 to -69 region of the human IL-8 gene (5'-agcttcacCAGTTGCAAATCGTGGAATTCCTctg-3' (SEQ

ID NO: 199) (binding sites for NF-IL6 and NF- $\kappa$ B are in boldface type) into HindIII-BamHI-cut TATA-CAT.

All transcription factor expression constructs used are described in Stein et al. (1993, Mol. Cell. Biol. 13:3964-3974). They are: pCMV4T-p65, which contains the cDNAs encoding human NF- $\kappa$ B p65; and plasmid pCMV4T-rC/EBPbeta, which contains a cDNA encoding the rat C/EBPbeta (also known as NF-IL6).

Transfection of cells and analysis of CAT activity are done as follows. All cell lines are cultured in Iscove's Dulbecco modified Eagle medium supplemented with 7.5% fetal calf serum and antibiotics. Mouse F9 embryonal carcinoma cells are transiently transfected by the calcium phosphate method (Graham et al., 1973, Virology 52:456-467) (also see *Current Protocols in Molecular Biology*, section 9.1.1, Supplement 14, 1990). Monkey COS cells are transiently transfected with plasmid DNA by the DEAE-dextran method (Kawai et al., 1984, Mol. Cell. Biol. 4:1172-1174) (see also *Current Protocols*, section 9.2.1, Supplement 14, 1987). Chloramphenicol-acetyltransferase (CAT) enzymatic activity was assayed as previously described using the Phase-Extraction Assay (see *Current Protocols in Molecular Biology*, Supplement 14, section 9.6.6), less preferably a chromatographic assay for CAT may be used (see *Current Protocols in Molecular Biology*, Supplement 14, section 9.6.3).

The following Tables 13 and 14 outline the different combinations of experiments that are used to verify the biological effect of each of the types of syngenes that are the subject of this specific embodiment of the invention. When examining these tables, it should be kept in mind that p65 is NF- $\kappa$ B and C/EBP is an NF-IL6.



Table 13

	Syngene Construct Used	Promoter - Transcription Factor Expressed	Promoter - Reporter Construct Used	Expected CAT Activity
5	None	pCMV4T-p65	pMHC-CAT	High
	"	pCMV4T-rC/EBPbeta	"	Low
	"	both **, see note below	"	Low
10	HSDB-NF $\kappa$ B-MHC	pCMV4T-p65	"	Low
	"	pCMV4T-rC/EBPbeta	"	Low
	"	both	"	Low
	HSDB-NF $\kappa$ B-IL-8	pCMV4T-p65	"	High
15	"	pCMV4T-rC/EBPbeta	"	Low
	"	both	"	Low
	HSDB-NF-IL-6	pCMV4T-p65	"	High
	"	pCMV4T-rC/EBPbeta	"	Low
20	"	both	"	Low
	None	pCMV4T-p65	pNF-IL-8-CAT	Med
	"	pCMV4T-rC/EBPbeta	"	Med
	"	both	"	High
25	HSDB-NF $\kappa$ B-MHC	pCMV4T-p65	"	Low
	"	pCMV4T-rC/EBPbeta	"	Med
	"	both	"	Med
	HSDB-NF $\kappa$ B-IL-8	pCMV4T-p65	"	Low
	"	pCMV4T-rC/EBPbeta	"	Med
30	"	both	"	Med
	HSDB-NF $\kappa$ B-IL-6	pCMV4T-p65	"	Med
	"	pCMV4T-rC/EBPbeta	"	Low
	"	both	"	Med

35

\*\* "both" refers to PMA treatment of cells, which leads to the activation of endogenous NF- $\kappa$ B and C/EBP and thus to the same results as co-transfection with both of these transcription factors.

Table 14

5	Syngene Construct Used	Promoter - Transcription Factor	Promoter - Reporter Construct Used	Expected CAT Activity
	None	pCMV4T-p65	pNF-IL-6-CAT	Med
	"	pCMV4T-rC/EBPbeta	"	Med
10	"	both	"	High
	HSDB-NF- $\kappa$ B-MHC	pCMV4T-p65	"	Med
	"	pCMV4T-rC/EBPbeta	"	Med
	"	both	"	High
15	HSDB-NF- $\kappa$ B-IL-8	pCMV4T-p65	"	Med
	"	pCMV4T-rC/EBPbeta	"	Med
	"	both	"	High
	HSDB-NF-IL-6	pCMV4T-p65	"	Med
20	"	pCMV4T-rC/EBPbeta	"	Low
	"	both	"	Med

It will be readily apparent from the results indicated in Tables 13 and 14 that the syngenes that bind to NF-IL6 and NF- $\kappa$ B sites are expected to selectively block the effect of the transcription factors NF-IL6 and NF- $\kappa$ B. In the case of the MHC-I gene reporter construct, only the HSDB directed toward the NF- $\kappa$ B site of the MHC gene regulates the MHC promoter element since that promoter has no NF-IL6 site and only one NF- $\kappa$ B site, specifically the one called NF- $\kappa$ B-MHC. The MHC gene promoter in this reporter is only activated by p65, the gene product of the NF- $\kappa$ B family gene known as Rel A. The MHC gene promoter has no NF-IL6 site and thus is not responsive to the co-expression of the NF-IL6 protein. Syngenes constructed with HSDB domains directed against

the NF- $\kappa$ B-MHC site can antagonize the activity of p65; but syngenes constructed with HSDB domains directed against other NF- $\kappa$ B sites have little or no effect on the transcription of the MHC reporter. This demonstrates one of the advantages of the use of syngenes. Syngenes have the capacity to discriminate between closely related binding sites of the same transcription factor.

Furthermore, it is apparent from the expected results in Tables 13 and 14, that those genes that have both NF- $\kappa$ B sites and NF-IL6 sites are partially responsive to the respective transcription factors and that the combination of p65 and NF-IL6 leads to synergistic activation of the IL-6 and IL-8 reporter gene constructs. It is also important to note that both of these reporter constructs are expected to respond poorly to the NF-IL6 protein when the syngene that expresses the HSDB directed against the NF- $\kappa$ B site is co-expressed. The synergism between the C/EBP $\beta$  and p65 protein is partially blocked by the HSDB-NF-IL-6 gene product or the HSDB-NF $\kappa$ B-IL6 gene product. However, the HSDB-NF $\kappa$ B-IL8 or HSDB-NF $\kappa$ B-MHC gene products should have little or no effect on the expression of the IL-6 reporter gene expression.

Thus, it is clear that the HSDB syngene products are expected to specifically and predictably suppress the expression of those gene promoters containing sites homologous to those of the respective target DNA fragments.

## 6.9. EXAMPLE: PREPARATION OF TSAR LIBRARIES

TSAR libraries have been prepared as set forth below.

### 6.9.1. PREPARATION OF THE TSAR-9 LIBRARY

#### 6.9.1.1. SYNTHESIS AND ASSEMBLY OF OLIGONUCLEOTIDES

Figure 5 shows the nucleotides and the assembly scheme used in construction of the TSAR-9 library. As

can be seen, the TSAR-9 library contains peptides with the amino acid sequence S(R/S)X<sub>1</sub>PGX<sub>1</sub>SR (SEQ ID NO: 200), in which X is an unpredictable amino acid. The oligonucleotides were synthesized with an applied Biosystems 380a synthesizer (Foster City, CA), and the full-length oligonucleotides were purified by HPLC.

Five micrograms of each of the pairs of oligonucleotides were mixed together in buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 0.001 % gelatin, 1.5 mM MgCl<sub>2</sub>) with 2 mM dNTP's, and 20 units of Taq DNA polymerase. The assembly reaction mixtures were incubated at 72°C for 30 seconds and then 30°C for 30 seconds; this cycle was repeated 60 times. It should be noted that the assembly reaction is not PCR, since a denaturation step was not used. Fill-in reactions were carried out in a thermal cycling device (Ericomp, LaJolla, CA) with the following protocol: 30 seconds at 72°C, 30 seconds at 30°C, repeated for 60 cycles. The lower temperature allows for annealing of the six base complementary region between the two sets of the oligonucleotide pairs. The reaction products were phenol/chloroform extracted and ethanol precipitated. Greater than 90% of the nucleotides were found to have been converted to double stranded synthetic oligonucleotides.

After resuspension in 300 µl of buffer containing 10 mM Tris-HCl, pH 7.5, 1 mM EDTA (TE buffer), the ends of the oligonucleotide fragments were cleaved with Xba I and Xho I (New England BioLabs, Beverly, MA) according to the supplier's recommendations. The fragments were purified by 4% agarose gel electrophoresis. The band of correct size was removed and electroeluted, concentrated by ethanol precipitation and resuspended in 100 µl TE buffer. Approximately 5% of the assembled oligonucleotides can be expected to have internal Xho I or Xba I sites; however, only the full-length molecules were used in the ligation step of the

assembly scheme. The concentration of the synthetic oligonucleotide fragments was estimated by comparing the intensity on an ethidium bromide stained gel run along with appropriate quantitated markers. All DNA

5 manipulations not described in detail were performed according to Sambrook, Fritsch and Maniatis, 1989, Molecular Cloning: A Laboratory Manual, 2d. ed. Cold Spring Harbor Laboratory Press.

To demonstrate that the assembled enzyme  
10 digested oligonucleotides could be ligated, the synthesized DNA fragments were examined for their ability to self-ligate. The digested fragments were incubated overnight at 18°C in ligation buffer with T4 DNA ligase. When the ligation products were examined by agarose gel  
15 electrophoresis, a concatamer of bands was visible upon ethidium bromide staining. As many as five different unit length concatamer bands (i.e., dimer, trimer, tetramer, pentamer, hexamer) were evident, suggesting that the synthesized DNA fragments were efficient  
20 substrates for ligation.

#### 6.9.1.2. CONSTRUCTION OF VECTORS

The construction of the M13 derived phage vectors useful for expressing a TSAR library has been  
25 recently described (Fowlkes et al., 1992, BioTechniques, 13:422-427). To express the TSAR-9 library, an M13 derived vector, m663, was constructed as described in Fowlkes et al. (id). Figure 6 illustrates the m663 vector containing the pIII gene having a c-myc-epitope,  
30 i.e., as a stuffer fragment, introduced at the mature N-terminal end, flanked by Xho I and Xba I restriction sites (see also, Figure 1 of Fowlkes et al., (id)).

#### 6.9.2. EXPRESSION OF THE TSAR-9 LIBRARY

35 The synthesized oligonucleotides were then ligated to Xho I and Xba I double-digested m663 RF DNA containing the pIII gene (Fowlkes, supra) by incubation

with ligase overnight at 12°C. More particularly, 50 ng of vector DNA and 5 ng of the digested synthesized DNA were mixed together in 50 µl ligation buffer (50 mM Tris, pH 8.0, 10 mM MgCl<sub>2</sub>, 20 mM DTT, 0.1 mM ATP) with T4 DNA  
5 ligase. After overnight ligation at 12°C, the DNA was concentrated by ethanol precipitation and washed with 70% ethanol. The ligated DNA was then introduced into *E. coli* (DH5αF'; GIBCO BRL, Gaithersburg, MD) by electroporation.

10 A small aliquot of the electroporated cells was plated and the number of plaques counted to determine that 10<sup>8</sup> recombinants were generated. The library of *E. coli* cells containing recombinant vectors was plated at a high density (~400,000 per 150 mM petri plate) for a  
15 single amplification of the recombinant phage. After 8 hr, the recombinant bacteriophage were recovered by washing each plate for 18 hr with SMG buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 0.05% gelatin) and after the addition of glycerol to 50% were frozen at  
20 -80°C. The TSAR-9 library thus formed had a working titer of ~2 x 10<sup>11</sup> pfu/ml.

#### 6.9.3. PREPARATION OF TSAR-12 LIBRARY

Figure 7 shows the formula for the synthetic  
25 oligonucleotides and the assembly scheme used in the construction of the TSAR-12 library. As can be seen, the TSAR-12 library contains peptides with the amino acid sequence S(S/T)X<sub>10</sub>φGδX<sub>10</sub>TR (SEQ ID NO: 201), in which X is an unpredictable amino acid, φ is S, R, G, C, or W, and δ  
30 is V, A, D, E, or G. As shown in Figure 7, the TSAR-12 library was prepared in substantially the same manner as the TSAR-9 library described in Section 6.9.1 and its subsections above with the following exceptions: (1) each of the variant non-predicted oligonucleotide  
35 sequences, i.e., NNB, was 30 nucleotides in length, rather than 54 nucleotides; (2) the restriction sites included at the 5' termini of the variant, non-predicted

sequences were Sal I and Spe I, rather than Xho I and Xba I; and (3) the invariant sequence at the 3' termini to aid annealing of the two strands was GCGGTG rather than CCAGGT (5' to 3').

5           After synthesis, including numerous rounds of annealing and chain extension in the presence of dNTP's and Taq DNA polymerase, and purification as described above in Section 6.9.1.1, the synthetic double stranded  
10           oligonucleotide fragments were digested with Sal I and Spe I restriction enzymes and ligated with T4 DNA ligase to the nucleotide sequence encoding the M13 pIII gene contained in the m663 vector to yield a library of  
15           TSAR-12 expression vectors as described in Section 6.9.3. The ligated DNA was then introduced into *E. coli* (DH5 $\alpha$ P'; GIBCO BRL, Gaithersburg, MD) by electroporation. The library of *E. coli* cells were plated at high density (~400,000 per 150 mm petri plate) for amplification of the recombinant phage. After about 8 hr, the recombinant bacteriophage were recovered by washing for 18 hr with  
20           SMG buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 0.05% gelatin) and after the addition of glycerol to 50% were frozen at -80°C.

          The TSAR-12 library thus formed had a working titer of ~ 2 x 10<sup>11</sup> pfu/ml.

25           The inserted synthetic oligonucleotides for each of the TSAR libraries, described in Section 6.9 above, had a potential coding complexity of 20<sup>36</sup> (~10<sup>47</sup>) and since ~10<sup>14</sup> molecules were used in each transformation experiment, each member of these TSAR libraries should be  
30           unique. After plate amplification the library solution or stock has 10<sup>4</sup> copies of each member/ml.

#### 6.9.4.     PREPARATION OF THE TSAR-13 AND TSAR-14 SEMIRIGID LIBRARIES

35           The following example illustrates yet another embodiment of a TSAR library expressing peptides that can form semirigid structures. The coding scheme which

encodes the variant residues in the oligonucleotides of this embodiment differs from that of the linear libraries described hereinabove.

5                    6.9.4.1. SYNTHESIS AND ASSEMBLY OF OLIGONUCLEOTIDES

Figure 8 shows the nucleotides used in the TSAR-13 and TSAR-14 libraries and the assembly scheme used in construction of the TSAR-13 library. The same oligonucleotide design was used for both TSAR-13 and  
10 TSAR-14 libraries. TSAR-13 was expressed in phagemid; TSAR-14 was expressed in phage. The oligonucleotides were designed to contain invariant nucleotides flanking contiguous sequences of unpredicted nucleotides. In this  
15 example, the single stranded nucleotide sequences when converted to double stranded oligonucleotides encode:  
(a) 5' to 3' Restriction site- cysteine, glycine - (NNK), - Gly-Cys-Gly- (NNK), - Complementary site Gly-Cys-Gly;  
and (b) 3' to 5' Complementary site Gly-Cys-Gly - (NNM), -  
20 Gly-Cys-Gly-Pro-Pro-Gly - Restriction site. Thus the library is designed to have semirigid binding domains each containing four cysteine residues that will form disulfide bonds in an oxidizing environment and adopt cloverleaf configurations. The additional proline  
25 residues were included to form a kink between the TSAR binding domain and the pIII or effector domain. In the design of the single stranded nucleotides, all 4 possible codons for glycine were utilized to help insure that the two single stranded nucleotides would anneal at the  
30 intended complementary glycine, cysteine, glycine encoding nucleotide sequence.

The oligonucleotides were synthesized with an Applied Biosystems 380a synthesizer (Foster City, CA) and the full length oligonucleotides were purified by gel  
35 electrophoresis.

To anneal the pairs of oligonucleotides, 200 pmol of each of the pair of oligonucleotides were mixed



together in Sequenase™ buffer (40 mM Tris pH 7.5, 20 mM MgCl<sub>2</sub>, 50 mM NaCl) with 0.1 ug/ml BSA, 10 mM DTT in a total volume of 200 ul. The mixture was incubated at 42°C for 5 minutes, then at 37°C for 15 minutes. Fill-in reactions were carried out by adding all four dNTPs to a concentration of 0.2 mM each and 20 units of Sequenase™, [(Version 2.0 (U.S. Biochemical, Cleveland, Ohio))] and incubating for 37°C for 15 minutes. Residual polymerase activity was heat inactivated by a 2 hour incubation at 65°C. After cooling, the ends of the oligonucleotide fragments were cleaved, by adding restriction buffer (10 mM Tris pH 7.5, 50 mM NaCl, 10 mM MgCl<sub>2</sub>), an additional 0.1 µg/ml BSA and an additional 2 mM DTT along with 300 units each of Xba I, Xho I. Three control reactions were run simultaneously. In the first control aliquot, i.e., a 10 µl aliquot of the fill in reaction, the first restriction enzyme (Xba I) was added to the same final concentration (units/µl). To the second control aliquot, the other restriction enzyme (Xho I) was added. No restriction enzyme was added to the third control aliquot. All samples were incubated for 2 hours at the temperature recommended by the restriction enzyme manufacturer. The cleaved oligonucleotides were extracted with an equal volume 1:1 phenol/chloroform, and ethanol precipitated. Fragments were purified on a 15% non-denaturing preparatory polyacrylamide gel in 1X TBE. The band of the correct size (as determined by comparison with control samples) was removed, isolated, ethanol precipitated and resuspended in TE buffer.

The recovered oligonucleotides can be inserted into an appropriate phage vector as described above, or into an appropriate phagemid vector, as described in Section 6.10.

#### 6.9.5. PREPARATION OF THE R26 LIBRARY

The R26 expression library was constructed essentially as described for the TSAR-9 library in

Section 6.9 and its subsections, except for the modifications depicted in Figure 9. The oligonucleotide assembly process depicted in Figure 9 results in expression of peptides with the following amino acid sequence:

$S(S/R)X_{12}\pi A\delta X_{12}SR$ , where  $\pi = S, P, T$  or  $A$ ; and  $\delta = V, A, D, E$  OR  $G$  (SEQ ID NO: 202)

#### 6.9.6. PREPARATION OF THE DC43 LIBRARY

The DC43 expression library was constructed essentially as described for the TSAR-9 library in Section 6.9.1 and its subsections, except for the modifications depicted in Figure 13. The oligonucleotide assembly process depicted in Figure 13 results in expression of peptides with the following amino acid sequence:

$HSS(S/R)X_{12}GCGX_{20}SRIEGRARPSR$  (SEQ ID NO: 203)

20

#### 6.10. CONSTRUCTION OF VECTOR pDAF1

The vector pDAF1 is constructed as follows:

To create the phagemid vector pDAF1, a segment of the M13 gene III was transferred into the Bluescript II SK+ vector (GenBank #52328). This vector replicates autonomously in bacteria, has an ampicillin drug resistance marker, and the fl origin of replication which allows the vector under certain conditions to be replicated and packaged into M13 particles. These M13 viral particles would carry both wild-type pIII molecules encoded by helper phage and recombinant pIII molecules encoded by the phagemid. These phagemids express only one to two copies of the recombinant pIII molecule and have been termed monovalent display systems (See Garrard et al., 1991, Biotechnol. 9:1373-1377). Rather than express the entire gene III, this vector has a truncated form of gene III [See generally, Lowman et al., 1991

(Biochemistry 30:10832-10838) which demonstrated that human growth hormone was more accessible to monoclonal antibodies when it was displayed at the NH<sub>2</sub>-terminus of a truncated form of pIII protein than at the NH<sub>2</sub>-terminus of the full-length form]. In the phagemid vector constructed here, the TSAR oligonucleotides are expressed at the mature terminus of a truncated pIII molecule, which corresponds to amino acids 198 to 406 of the mature pIII molecules.

10 The preferred vector is pDAF, which encodes amino acids 198-406 of the pIII protein, a short polylinker within the pIII gene and the linker gly-gly-gly-ser between the polylinker and the pIII molecule. This plasmid expresses pIII from the promoter and  
15 utilizes the PelB leader sequence for direction of pIII's compartmentalization to the bacterial membrane for proper M13 viral assembly.

A pair of oligonucleotides were designed  
CGTTACGAATTCTTAAGACTCCTTATTACGCA (SEQ ID NO: 204) and  
20 CGTTAGGATCCCCATTCGTTTCTGAATATCAA (SEQ ID NO: 205) to amplify a portion (aa 198-406) of the pIII gene from M13mp8 DNA via PCR. Since these oligonucleotides carried Bam HI and Eco RI sites near the 5' termini, the PCR product was then digested with Bam HI and Eco RI, ligated  
25 with pBluescript II SK+ DNA digested with the same enzymes, and introduced into *E. coli* by transformation. After the recombinant was identified, an additional double-stranded DNA segment was cloned into it, encoding the PelB signal leader with an upstream ribosome binding  
30 site. This segment was prepared by PCR from *E. coli* DNA using the oligonucleotides  
GCCGACGCGACGAGCTCGACTGCAAATTCTATTTCAA (SEQ ID NO: 206) and  
CTAATGTCTAGAAAGCTTCTCGAGCCCTGCAGCTGCACCTGGGCCATCGACTGG  
(SEQ ID NO: 207). The termini of the PCR product  
35 introduced a short polylinker of Pst I, Xho I, Hind III, and Xba I sites into the vector. The Xho I and Xba I sites were positioned so that assembled TSAR

oligonucleotides could be cloned and expressed in the same reading frame as in the phage vectors described above. The third and final segment of DNA introduced into the vector, encoded the linker sequence GGGGS (SEQ ID NO: 56) between the polylinker and gene III. This linker matches a repeated sequence motif of the pIII molecule and was included in the chimeric gene to create a swivel point separating the expressed peptide and the pIII protein molecule. This vector has been named pDAF1. Figure 10A schematically illustrates the pDAF1 phagemid vector.

#### 6.10.1. CONSTRUCTION OF VECTORS pDAF2 AND pDAF3

The vectors pDAF2 and pDAF3 are prepared from pDAF1 but differ from the parent vector in that each contains the c-myc encoding sequence at the NH<sub>2</sub> and COOH terminal sides, respectively, of the polylinker of Pst I, Xho I, Hind III and Xba I restriction sites. Figure 10B and C schematically illustrate the phagemid vectors pDAF2 and pDAF3. The pDAF2 and pDAF3 vectors are constructed as shown schematically in Figure 10D.

#### 6.10.2. CONSTRUCTION OF THE PHAGEMID VECTORS

The construction of the phagemid vector pDAF1 is described in Section 6.10. This vector was modified to include the full length pIII gene by inserting the amino-terminus of the pIII gene from m666. Both pDAF3 and m666 were cut with AlwN I, and Xba I, and a 0.7 kb fragment was transferred from m666 to pDAF3 to generate the vector pFLP3.

#### 6.11. EXPRESSION OF THE TSAR-13 PHAGEMID LIBRARY

The synthesized oligonucleotides were ligated to Xba I, Xho I double-digested pFLP3 DNA, electroporated into XL1 blue *E. coli*. An aliquot was plated and the titer was determined to be  $8 \times 10^7$  total colonies. The entire library was plated on ampicillin plates. To

express the TSAR-13 library,  $7 \times 10^{10}$  cells were added to 30 ml of 2xYT media and incubated for 30 min at 37°C, after which  $4 \times 10^{11}$  pfu of M13K07 helper phage were added. Aliquots were induced by adding 0.004% IPTG, 2% glucose or nothing and incubated for 1 hr. Then 150 ml of 2xYT media plus 70 µg/ml of kanamycin were added and the cells were further incubated for 4 hr at 37°C. Phagemid particles were PEG precipitated, collected by centrifugation and resuspended in 4 ml of media. The titer of each was  $2 \times 10^{12}$  pfu/ml. The total number of recombinants was  $8 \times 10^7$ .

The inserted synthetic oligonucleotides for the TSAR-13 library had a potential coding complexity of  $20^{24}$  ( $1.68 \times 10^{31}$ ) and since  $10^{12}$  molecules were used in each transformation experiment, each member of the library should be unique.

#### 6.12. CONSTRUCTION OF PHAGE VECTORS

To express the TSAR-14 library, a member of the TSAR-9 library, as described in Section 6.9.1 and its subsections, above, was modified by cutting out the polylinker using Eco RI and Hind III and inserting pUC18 polylinker (previously modified by deleting the Xba I site) to produce blue plaques.

##### 6.12.1. EXPRESSION OF THE TSAR-14 PHAGE LIBRARY

The synthesized oligonucleotides were then ligated to Xba I, Xho I double digested m666 containing the pIII gene as described for the TSAR-9 library. The ligated DNA was then introduced into *E. coli* cells by electroporation.

#### 6.13. PREPARATION OF THE R8C LIBRARY

A random peptide expression library, termed R8C was prepared as depicted in Figure 11. The oligonucleotide assembly process depicted in Figure 11 predominantly yields an expressed peptide with a random

8-mer sequence, comprising the following amino acid sequence expressed at the amino terminus of pIII: SSC(X),CGSR (SEQ ID NO: 208). However, a percentage of the library contains a double insert resulting in expression of a peptide with a random 16-mer sequence (see Figure 12), comprising the following amino acid sequence expressed at the amino terminus of pIII: SSC(X),CGSRST(X),TTR...(SEQ ID NO: 209).

#### 10           6.14.   IDENTIFICATION OF LIGAND BINDING TSARS

In several series of experiments, the TSAR-9, TSAR-12, TSAR-13, TSAR-14, R8C, and R26 libraries described above were screened for expressed proteins/peptides having binding specificity for a variety of different ligands of choice.

##### 6.14.1.   METHODS FOR SCREENING

The following methods were employed to screen the libraries, except as otherwise noted.

20           The ligand of choice was conjugated to magnetic beads, obtained from one of two sources: Amine Terminated particulate supports, #8-4100B (Advanced Magnetix, Cambridge, MA) and Dynabeads M-450, tosylactivated (Dynal, Great Neck, NY), according to the instructions of the manufacturer. To block any unreacted groups and non-specific binding to the beads, the beads were incubated with excess bovine serum albumin (BSA). The beads were then washed with numerous cycles of suspension in PBS-0.05% Tween ® 20, and recovered with a strong magnet. The beads were then stored at 4°C until needed.

35           In the screening experiments, 1 ml of library was mixed with 100 µl of resuspended beads (1-5 mg/ml). The tube contents were tumbled at 4°C for 1-2 hrs. The magnetic beads were then recovered with a strong magnet and the liquid was removed by aspiration. The beads were then washed by adding 1 ml of PBS-0.05% Tween ® 20,

inverting the tube several times to resuspend the beads, drawing the beads to the tube wall with the magnet and moving the liquid contents. The beads were washed repeatedly 5-10 additional times. Fifty  $\mu$ l of 50 mM glycine-HCl (pH 2.2), 100 mg/ml BSA solution were added to the washed beads to denature proteins and release bound phage. After 5-10 minutes, the beads were pulled to the side of the tubes with a strong magnet and the liquid contents then transferred to clean tubes. To the tubes, 100  $\mu$ l 1 M Tris-HCl (pH 7.5) or 1 M  $\text{NaH}_2\text{PO}_4$  (pH 7) was added to neutralize the pH of the phage sample. The phage were then serially diluted from  $10^{-3}$  to  $10^{-6}$  and aliquots plated with *E. coli* DH5 $\alpha$ F' cells to determine the number of plaque forming units of the sample. In certain cases, the platings were done in the presence of XGal and IPTG for color discrimination of plaques (i.e., lacZ' plaques are blue, lacZ- plaques are white). The titer of the input samples was also determined for comparison (dilutions were generally  $10^{-6}$  to  $10^{-9}$ ).

Successful screening experiments have generally involved 3 rounds of serial screening conducted in the following manner. First, the library was screened and the recovered phage rescreened immediately. Second, the phage that were recovered after the second round were plate amplified, according to Maniatis. The phage were eluted into SMG buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 10 mM  $\text{MgCl}_2$ , 0.05% gelatin), by overlaying the plates with ~5 ml of SMG buffer and incubating the plates at 4°C overnight. Third, a small aliquot was then taken from the plate and rescreened. The recovered phage were then plated at a low density to yield isolated plaques for individual analysis.

The individual plaques were picked with a toothpick and used to inoculate cultures of *E. coli* F cells in 2xYT. After overnight culture at 37°C, the cultures were then spun down by centrifugation. The liquid supernatant was then transferred to a clean tube

and served as the phage stock. Generally, it has a titer of  $10^{12}$  pfu/ml which is stable at 4°C. Individual phage aliquots were then retested for their binding to the ligand coated beads and their lack of binding to other control beads (i.e., BSA coated beads, or beads conjugated with other ligand).

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.



WHAT IS CLAIMED IS:

1. A pharmaceutical composition comprising:
- 5 (a) a nucleic acid encoding a protein, said protein comprising a binding domain which binds to a ligand of choice, in which the nucleotide sequence encoding said binding domain is a sequence identified by a method comprising screening a library of
- 10 recombinant vectors, said vectors comprising unpredictable nucleotides arranged in one or more contiguous sequences; and
- 15 (b) a pharmaceutically acceptable carrier.
2. The composition of claim 1 where the total number of unpredictable nucleotides is greater than or equal to about 15 and less than or equal to about 600.
- 20 3. A pharmaceutical composition comprising:
- (a) a nucleic acid encoding a protein, said protein comprising a binding domain which binds to a ligand of choice, in which the nucleotide sequence encoding said binding
- 25 domain is a sequence identified by a method comprising screening a library of recombinant vectors, said vectors encoding a plurality of heterofunctional fusion polypeptides, said fusion polypeptides
- 30 comprising:
- (i) a binding region encoded by a first oligonucleotide comprising unpredictable nucleotides in which the unpredictable
- 35 nucleotides are arranged in one or more contiguous sequences; and

- 5 (ii) an effector domain encoded by a second oligonucleotide, said effector domain being a peptide that enhances expression or detection of the binding region; and
- (b) a pharmaceutically acceptable carrier.
- 10 4. A pharmaceutical composition comprising:
- (a) a nucleic acid encoding a protein, said protein comprising a binding domain which binds to a ligand of choice, in which the nucleotide sequence encoding said binding domain is a sequence identified by a
- 15 method comprising screening a library of recombinant vectors, said vectors encoding a plurality of heterofunctional fusion polypeptides, said fusion polypeptides comprising:
- 20 (i) a binding region encoded by a first oligonucleotide comprising unpredictable nucleotides in which the unpredictable nucleotides are arranged in one
- 25 or more contiguous sequences and the contiguous sequences are flanked by invariant residues designed to encode amino acids that confer a
- 30 desired structure to the binding region; and
- (ii) an effector domain encoded by a second oligonucleotide, said effector domain being a peptide that enhances expression or
- 35 detection of the binding region; and

(b) a pharmaceutically acceptable carrier.

5. A pharmaceutical composition comprising:

- 5 (a) a nucleic acid encoding a protein,  
said protein comprising a binding  
domain which binds to a ligand of  
choice, in which the amino acid  
sequence of said binding domain is  
10 identified by a method comprising  
screening a chemically synthesized  
peptide library, in which the  
peptides of said library comprise one  
or more contiguous sequences of  
unpredictable amino acids, wherein  
15 the total number of unpredictable  
amino acids is greater than or equal  
to 5 and less than or equal to 25;  
and  
(b) a pharmaceutically acceptable  
20 carrier.

6. The pharmaceutical composition of claim 1  
in which the coding strand of the unpredictable  
nucleotides comprises the formula  $(NNB)_n \dots$  where

- 25 N is A, C, G or T;  
B is G, T or C; and  
n and m are integers, such that  
 $20 \leq n + m \leq 200$ .

30 7. The composition of claim 1 in which said  
nucleic acid is at least part of an expression vector  
which expresses said nucleic acid in a suitable host  
cell.

35 8. The composition of claim 1 in which said  
nucleic acid comprises terminal sequences at each end

which promote homologous recombination with genomic sequences.

9. The composition of claim 1 in which said  
5 protein further comprises a nuclear localization signal.

10. The composition of claim 1 in which said  
protein further comprises a transcriptional activation  
signal.

10

11. The composition of claim 1, in which said  
ligand is selected from the group consisting of a non-  
ionic chemical group, an ion, a metal, a protein or  
portion thereof, a peptide or portion thereof, a nucleic  
15 acid or portion thereof, a carbohydrate, a lipid, a viral  
particle or portion thereof, a membrane vesicle or  
portion thereof, a cell wall component, a synthetic  
organic compound, a bioorganic compound and an inorganic  
compound.

20

12. The composition of claim 1, in which said  
ligand is a ligand which binds to a naturally occurring  
receptor selected from the group consisting of the  
variable region of an antibody, an enzyme/substrate  
25 binding site, an enzyme/co-factor binding site, a  
regulatory DNA binding protein, an RNA binding protein, a  
binding site of a metal binding protein, a nucleotide  
fold or GTP binding protein, a calcium binding protein, a  
membrane protein, a viral protein and an integrin.

30

13. The composition of claim 1 in which said  
ligand is selected from the group consisting of a  
molecule comprising a transcriptional regulatory site on  
DNA; a transcriptional regulator that binds to a  
35 transcriptional regulatory site on DNA; and a first  
protein that binds to a second protein, said second

protein being a transcriptional regulator that binds to a transcriptional regulatory site on DNA.

14. The composition of claim 13 in which the  
5 ligand comprises a transcriptional regulatory site on DNA.

15. The composition of claim 14 in which the  
ligand comprises a sequence selected from the group  
10 consisting of: the sequence 5'GGGTGGGGATTCCCCATCT3' (SEQ  
ID NO: 135), the sequence 5'ATGTGGGATTTTCCCATG3' (SEQ ID  
NO: 137), the sequence 5'ATCGTGGGAATTTCTCTG3' (SEQ ID NO:  
139), and the sequence 5'ACGTCATTGCACAATCTT3' (SEQ ID NO:  
141).

15

16. The composition of claim 14 in which the  
transcriptional regulatory site on DNA is an NF- $\kappa$ B  
nucleic acid binding site.

20

17. The composition of claim 16 in which said  
binding domain binds to an H2 $\kappa$ B nucleic acid binding  
site, but does not substantially bind to an IL-6 $\kappa$ B or to  
an IL-8 $\kappa$ B nucleic acid binding site.

25

18. The composition of claim 16 in which said  
binding domain binds to an IL-6 $\kappa$ B, IL-8 $\kappa$ B, and H2 $\kappa$ B  
nucleic acid binding site.

19. The composition of claim 14 in which the  
30 transcriptional regulatory site on DNA is selected from  
the group consisting of a GATA transcription factor  
nucleic acid binding site, an AP-1 nucleic acid binding  
site, and an ATF nucleic acid binding site.

35

20. The composition of claim 13 in which said  
ligand is NF- $\kappa$ B.

21. The composition of claim 13 in which said ligand is IF- $\kappa$ B.

22. A pharmaceutical composition comprising:

- 5 (a) a protein comprising a binding domain which binds to a ligand of choice, in which the amino acid sequence encoding said binding domain is an amino acid sequence identified by a method comprising  
10 screening a library of recombinant vectors, said vectors comprising a nucleotide sequence encoding unpredictable amino acids arranged in one or more contiguous sequences, wherein the total  
15 number of unpredictable amino acids is greater than or equal to 20 and less than or equal to about 200, in which said ligand is selected from the group consisting of an NF- $\kappa$ B nucleic acid binding site, NF- $\kappa$ B, and IF- $\kappa$ B; and  
20 (b) a pharmaceutically acceptable carrier.

23. A nucleic acid encoding a protein, said protein comprising a binding domain which binds to a  
25 ligand of choice, in which the nucleotide sequence encoding said binding domain is a sequence identified by a method comprising screening a library of recombinant vectors, said vectors comprising unpredictable nucleotides arranged in one or more contiguous sequences,  
30 wherein the total number of unpredictable nucleotides is greater than or equal to 15 and less than or equal to about 600 and in which said protein further comprises a sequence providing for *in vivo* or intracellular targeting of said protein.

35

24. The nucleic acid of claim 23 in which said protein further comprises a transcriptional activation sequence.

5           25. The nucleic acid of claim 23 in which said nucleotide sequence is flanked by sequences which promote homologous recombination with genomic sequences.

10           26. The nucleic acid of claim 23 in which the coding strand of the unpredictable nucleotides comprises the formula  $(NNB)_n \dots_m$  where

N is A, C, G or T;

B is G, T or C; and

n and m are integers, such that

15            $20 \leq n + m \leq 200$ .

27. The nucleic acid of claim 23 in which said sequence providing for targeting is a nuclear localization sequence.

20

28. The nucleic acid of claim 23 in which said ligand is selected from the group consisting of a molecule comprising a transcriptional regulatory site on DNA; a transcriptional regulator that binds to a transcriptional regulatory site on DNA; and a first protein that binds to a second protein, said second protein being a transcriptional regulator that binds to a transcriptional regulatory site on DNA.

30           29. The nucleic acid of claim 28 in which the ligand comprises a transcriptional regulatory site on DNA.

35           30. The nucleic acid of claim 29 in which the ligand comprises a sequence selected from the group consisting of: the sequence 5'GGGTGGGGATTCCCCATCT3' (SEQ ID NO: 135), the sequence 5'ATGTGGGATTTTCCCATG3' (SEQ

NO: 137), the sequence 5'ATCGTGGAATTTTCCTCTG3' (SEQ ID NO: 139), and the sequence 5'ACGTCATTGCACAATCTT3' (SEQ ID NO: 141).

5           31. The nucleic acid of claim 29 in which the transcriptional regulatory site on DNA is an NF- $\kappa$ B nucleic acid binding site.

10           32. The nucleic acid of claim 31 in which said binding domain binds to an H2 $\kappa$ B nucleic acid binding site, but does not substantially bind to an IL-6 $\kappa$ B or to an IL-8 $\kappa$ B nucleic acid binding site.

15           33. The nucleic acid of claim 31 in which said binding domain binds to an IL-6 $\kappa$ B, IL-8 $\kappa$ B, and H2 $\kappa$ B nucleic acid binding site.

20           34. The nucleic acid of claim 29 in which the transcriptional regulatory site on DNA is selected from the group consisting of a GATA transcription factor nucleic acid binding site, an AP-1 nucleic acid binding site, and an ATF nucleic acid binding site.

25           35. The nucleic acid of claim 23 in which said ligand is NF- $\kappa$ B.

          36. The nucleic acid of claim 23 in which said ligand is IF- $\kappa$ B.

30           37. A method of modifying transcription of one or more genes of interest comprising delivering to a cell which is capable of expressing one or more genes of interest a composition comprising a nucleic acid encoding a protein, said protein comprising a binding domain which  
35 binds to a ligand of choice, in which the nucleotide sequence encoding said binding domain is a sequence identified by a method comprising screening a library of



recombinant vectors, said vectors comprising unpredictable nucleotides arranged in one or more contiguous sequences, wherein the total number of unpredictable nucleotides is greater than or equal to 15 and less than or equal to about 600; in which said ligand is selected from the group consisting of a molecule comprising a transcriptional regulatory site on DNA, a DNA binding protein that is a transcriptional regulator, and a protein that binds to said DNA binding protein.

10

38. The method of claim 37 which is carried out *in vitro*.

39. The method of claim 37 which is carried out *in vivo*, in which said delivering is carried out by administering said composition to a subject.

40. The method of claim 37 in which said ligand comprises a transcriptional regulatory site on DNA.

41. The method of claim 40 in which the ligand comprises a sequence selected from the group consisting of: the sequence 5'GGGTGGGGATTCCCCATCT3' (SEQ ID NO: 135), the sequence 5'ATGTGGGATTTTCCCATG3' (SEQ ID NO: 137), the sequence 5'ATCGTGGGAATTTCTCTG3' (SEQ ID NO: 139), and the sequence 5'ACGTCATTGCACAATCTT3' (SEQ ID NO: 141).

42. The method of claim 40 in which the transcriptional regulatory site on DNA is an NF- $\kappa$ B nucleic acid binding site.

43. The method of claim 42 in which said binding domain binds to an H2 $\kappa$ B nucleic acid binding site, but does not substantially bind to an IL-6 $\kappa$ B or to an IL-8 $\kappa$ B nucleic acid binding site.

44. The method of claim 42 in which said binding domain binds to an IL-6κB, IL-8κB, and H2κB nucleic acid binding site .

5 45. The method of claim 40 in which the transcriptional regulatory site on DNA is selected from the group consisting of a GATA transcription factor nucleic acid binding site, an AP-1 nucleic acid binding site, and an ATF nucleic acid binding site.

10

46. The method of claim 37 in which said ligand is a protein that is a transcriptional regulator that binds to a transcriptional regulatory site on DNA.

15

47. The method of claim 37 in which said ligand is a first protein that binds to a second protein, said second protein being a transcriptional regulator that binds to a transcriptional regulatory site on DNA.

20

48. The method of claim 37 in which said protein, when expressed in a suitable cell, inhibits transcription of one or more genes of interest.

25

49. The method of claim 37 in which said protein, when expressed in a suitable cell, increases transcription of one or more genes of interest.

30

50. The method of claim 46 in which said transcriptional regulator is NF-κB.

51. The method of claim 47 in which said ligand is IF-κB.

35

52. The method of claim 37 in which said protein further comprises a nuclear localization signal.

53. The method of claim 37 in which said protein further comprises a transcriptional activation signal.

5 54. A method of modifying transcription of one or more genes of interest comprising delivering to a cell which is capable of expressing one or more genes of interest a composition comprising a nucleic acid encoding a protein, said protein comprising a binding domain which  
10 binds to a ligand of choice, in which the amino acid sequence of said binding domain is identified by a method comprising screening a chemically synthesized peptide library, in which the peptides of said library comprise one or more contiguous sequences of unpredictable amino  
15 acids, wherein the total number of unpredictable amino acids is greater than or equal to 5 and less than or equal to 25; in which said ligand is selected from the group consisting of a molecule comprising a transcriptional regulatory site on DNA, a DNA binding  
20 protein that is a transcriptional regulator, or a protein that binds to said DNA binding protein.

55. A therapeutic method comprising administering to a subject a therapeutically effective  
25 amount of any one of the pharmaceutical compositions of claims 1-22.

56. A method for identifying a nucleic acid that encodes a peptide which binds to a ligand of choice,  
30 comprising screening a library of recombinant vectors which express a plurality of proteins comprising a binding domain encoded by an oligonucleotide, said oligonucleotide comprising unpredictable nucleotides, in which the unpredictable nucleotides are arranged in one  
35 or more contiguous sequences, wherein the total number of unpredictable nucleotides is greater than or equal to about 15 and less than or equal to about 600; in which

said screening is done by a method comprising (a) contacting the plurality of proteins with said ligand of choice under conditions conducive to ligand binding, in which the ligand of choice is selected from the group consisting of a molecule comprising a transcriptional regulatory site on DNA, a DNA binding protein that is a transcriptional regulator, and a protein that binds to said DNA binding protein.

10                   57. The method of claim 56 in which said ligand comprises a transcriptional regulatory site on DNA.

15                   58. The method of claim 57 in which the ligand comprises a sequence selected from the group consisting of: the sequence 5'GGGTGGGGATTCCCCATCT3' (SEQ ID NO: 135), the sequence 5'ATGTGGGATTTTCCCATG3' (SEQ ID NO: 137), the sequence 5'ATCGTGGAATTTCTCTG3' (SEQ ID NO: 139), and the sequence 5'ACGTCATTGCACAATCTT3' (SEQ ID NO: 20   141).

25                   59. The method of claim 57 in which the transcriptional regulatory site on DNA is an NF- $\kappa$ B nucleic acid binding site.

30                   60. The method of claim 59 in which said binding domain binds to an H2 $\kappa$ B nucleic acid binding site, but does not substantially bind to an IL-6 $\kappa$ B or to an IL-8 $\kappa$ B nucleic acid binding site.

35                   61. The method of claim 59 in which said binding domain binds to an IL-6 $\kappa$ B, IL-8 $\kappa$ B, and H2 $\kappa$ B nucleic acid binding site.

40                   62. The method of claim 57 in which the transcriptional regulatory site on DNA is selected from the group consisting of a GATA transcription factor

nucleic acid binding site, an AP-1 nucleic acid binding site, and an ATF nucleic acid binding site.

63. The method of claim 56 in which said  
5 ligand is a DNA binding protein that is a transcriptional regulator.

64. The method of claim 56 in which said  
10 ligand is a protein that binds to said DNA binding protein.

65. A method for identifying a peptide which  
binds to a ligand of choice, comprising screening a  
chemically synthesized peptide library, in which the  
15 peptides of said library comprise one or more contiguous  
sequences of unpredictable amino acids, wherein the total  
number of unpredictable amino acids is greater than or  
equal to 5 and less than or equal to 25; in which said  
screening is done by a method comprising (i) contacting  
20 the plurality of peptides with said ligand of choice  
under conditions conducive to ligand binding, in which  
the ligand of choice is selected from the group  
consisting of a molecule comprising a transcriptional  
regulatory site on DNA, a DNA binding protein that is a  
25 transcriptional regulator, and a protein that binds to  
said DNA binding protein, and (ii) recovering a peptide  
which binds the ligand of choice.

66. The method according to claim 56 in which  
30 the coding strand of the unpredictable nucleotides  
comprises the formula  $(NNB)_n..._m$  where

N is A, C, G or T;

B is G, T or C; and

n and m are integers, such that

35  $20 \leq n + m \leq 200$ .

67. The composition of claim 1 or 2 which further comprises a promoter operably linked to the nucleic acid.

5                   68. A protein which binds to an NF- $\kappa$ B nucleic acid binding site, encoded by a nucleic acid identified by the method of claim 56.

10                   69. A method of modifying the activity of one or more genes of interest comprising delivering to a cell which is capable of expressing one or more genes of interest a composition comprising a nucleic acid encoding a protein, said protein comprising a binding domain which binds to (a) said one or more genes of interest, or (b) a  
15                   protein product encoded by said one or more genes of interest, in which the nucleotide sequence encoding said binding domain is a sequence identified by a method comprising screening a random peptide library.

20                   70. A molecule comprising a peptide having an amino acid sequence selected from the group consisting of:

SWCTYSGYCRVSSAGTAQRTSVDRDGM (SEQ ID NO: 143);  
RTGNEQPPGSFGRAGCFHPGCKYMKLN (SEQ ID NO: 144);  
25                   SDKYFHDIRKYHPAAATSYKTRPDMPST (SEQ ID NO: 145);  
REWGVPGAHNRIRDHCNGPRCHAIRTNASHTQHISRPPD (SEQ ID NO: 146);  
SIDKVQPPGTSGRKTGCFHPGCKYIKLK (SEQ ID NO: 147);  
STGKEHPPGSFGRATGCFHPGCKYIKLK (SEQ ID NO: 148);  
STGKEHSPGSLGRAPGCWHPGCKYIKLK (SEQ ID NO: 149);  
30                   RTGNDQPPRSFGHATGCFHPGCKYIKNK (SEQ ID NO: 150);  
STGKDHPSSFGRAAGCFHPGCKYIKHK (SEQ ID NO: 151);  
RTGEDHPPGSFGKAAGCFHPGCKYIKHK (SEQ ID NO: 152);  
RTGNDHPSVSYGRASGCFHAGCKYIKHK (SEQ ID NO: 153);  
RTGNENPPGSLGRGTGCFHAACKYIKHK (SEQ ID NO: 154);  
35                   RAGTDQPPGSFVRASGCFHPGCKYIKHK (SEQ ID NO: 155);  
RTGVDQSPESFDRATGCFHPGCKYIKHN (SEQ ID NO: 156);  
RTGNDDPTGALARNSGCYHPGCKYINHK (SEQ ID NO: 157);

RTGNDNPPGSSGPASGCFHPGCKYMKHK (SEQ ID NO: 158);  
 STGKEHPPGSFGRVSGCFHPGCKYMKHN (SEQ ID NO: 159);  
 RTGNDHPSGSFEHAAGCFHPGCKYMKHK (SEQ ID NO: 160);  
 RTGDDHPTGSIGRASGCFHPGCKYIKHN (SEQ ID NO: 161);  
 5 STGKEHPPGSIGRASGCFHPGCKYIRHT (SEQ ID NO: 162);  
 RTGNEHPPGSIGRASGCFHPGCNYIKHK (SEQ ID NO: 163);  
 RTGNEHPPGSLGRAPGCYHPGCKYIHHP (SEQ ID NO: 164);  
 RTGNYDPPGSFPGSTGCFHPGCKYIHHK (SEQ ID NO: 165);  
 STGNEQHPSGLRRASGCFHPGCKYIHHN (SEQ ID NO: 166);  
 10 STGNEHPPGSVGHAPGCFHPGCKYMLK (SEQ ID NO: 167);  
 SADLVAPSGENRSGRGCFHPGCKYIKHKPTTPPAPSSA (SEQ ID NO: 168);  
 SAATADLVGERSGRGCFHPGCKYIKHKGSTSQSQPL (SEQ ID NO: 169);  
 SPDTKSPGGDVRSGRGCFHPGCKYIKHKSDDSHKEATT (SEQ ID NO: 170);  
 SVTSDDPQGRERSGRGCFHPGCKYIKHKNHATSESPNL (SEQ ID NO: 171);  
 15 SGREPEGPSEQRSGRGCFHPGCKYIKHKTQSTMHTTPA (SEQ ID NO: 172);  
 SFRALDEPTATRSRGCFHPGCKYIKHKSPKDSFPNTTAA (SEQ ID NO: 173);  
 SGTVPTEAVNNRPGCFHPGCKYIKHKLPRQPSPTPPLA (SEQ ID NO: 174);  
 SDWSERYGGERGGGCFHPGCKYIKHKNKNSGTPLPSDP (SEQ ID NO: 175);  
 SNTDTAWAGNGDRGCFHPGCKYIKHKTTPPTNKTYETP (SEQ ID NO: 176);  
 20 SGNVDNGTERGGKGCFFHPGCKYIKHKRPNAYEVVPLD (SEQ ID NO: 177);  
 SERAERGEWNWSAGCFHPGCKYIKHKSPRGANRSLVGA (SEQ ID NO: 178);  
 and  
 SSDGDTPPGSRKGGCFHPGCKYIKHKPSPQLPREGTHN (SEQ ID NO: 179);  
 or a binding portion thereof.

25

71. A pharmaceutical composition comprising the molecule of claim 70 and a suitable pharmaceutical carrier.

30

72. A method of modifying the activity of one or more genes of interest comprising delivering to a cell which is capable of expressing one or more genes of interest a composition comprising a nucleic acid encoding a protein, said protein comprising a binding domain which  
 35 binds to (i) said one or more genes of interest, or (ii) a protein product encoded by said one or more genes of interest, in which the nucleotide sequence encoding said

binding domain is a sequence identified by a method comprising:

- (a) screening a random peptide library to identify a preliminary binding domain;
- 5 (b) subjecting the preliminary binding domain to a process of directed evolution to identify said binding domain.

73. The method of claim 74 where said binding  
10 domain specifically binds an H2κB site.

15

20

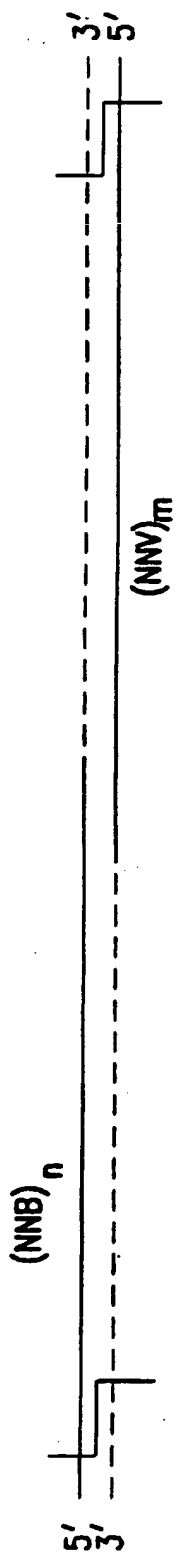
25

30

35



1/24



COMPLEMENTARY REGION

ANNEAL TWO OLIGOS,  
EXTEND WITH DNA POLYMERASE,  
DIGEST & LIGATE TO VECTOR

FIG.1A

2/24

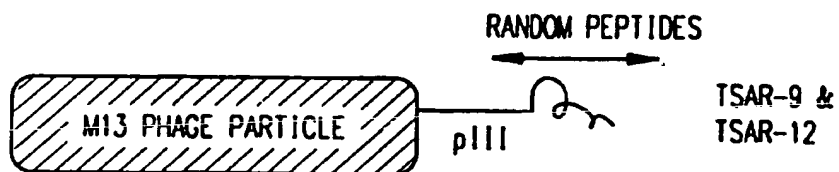


FIG.1B

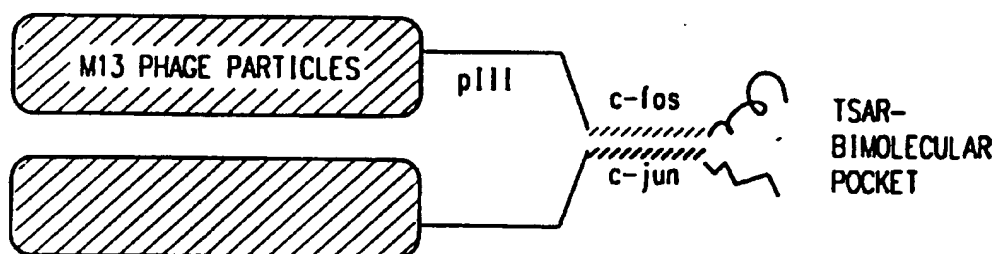


FIG.1C

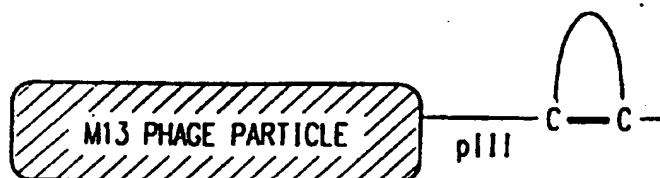


FIG.1D

3/24

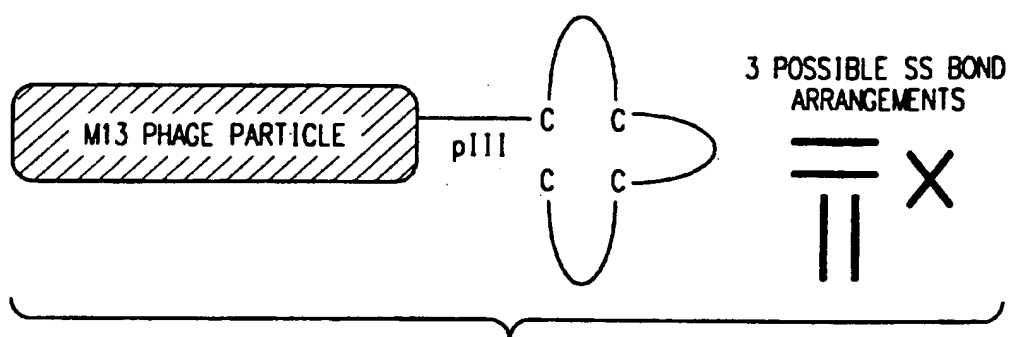


FIG.1E

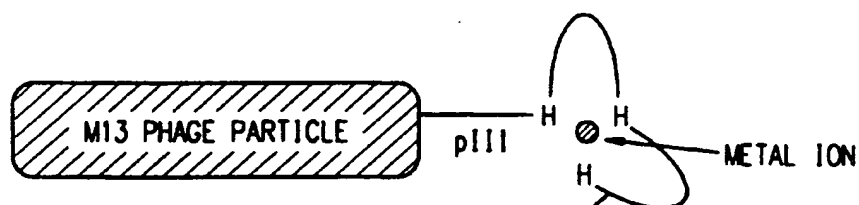


FIG.1F

4/24

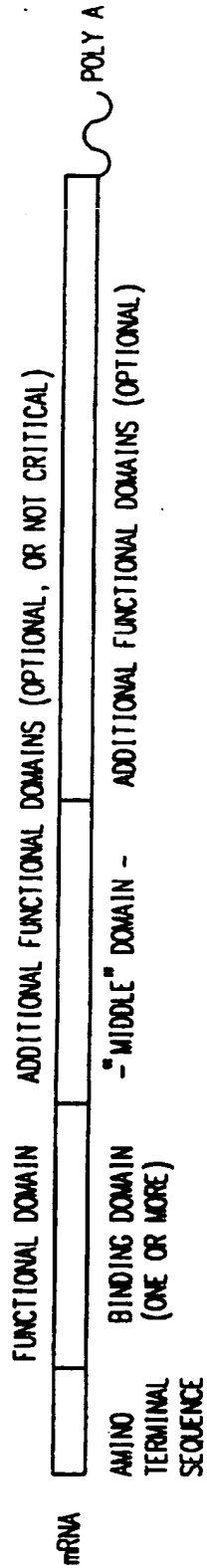


FIG.2

5/24

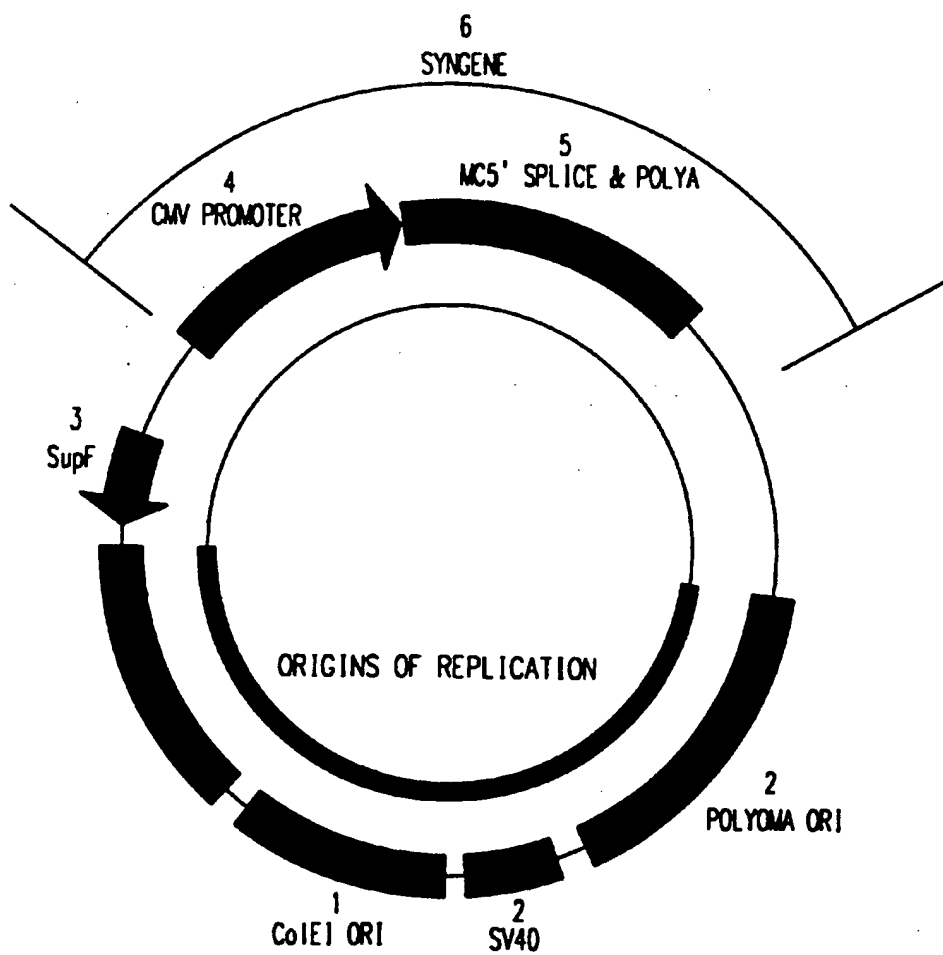


FIG.3

SUBSTITUTE SHEET (RULE 26)

6/24

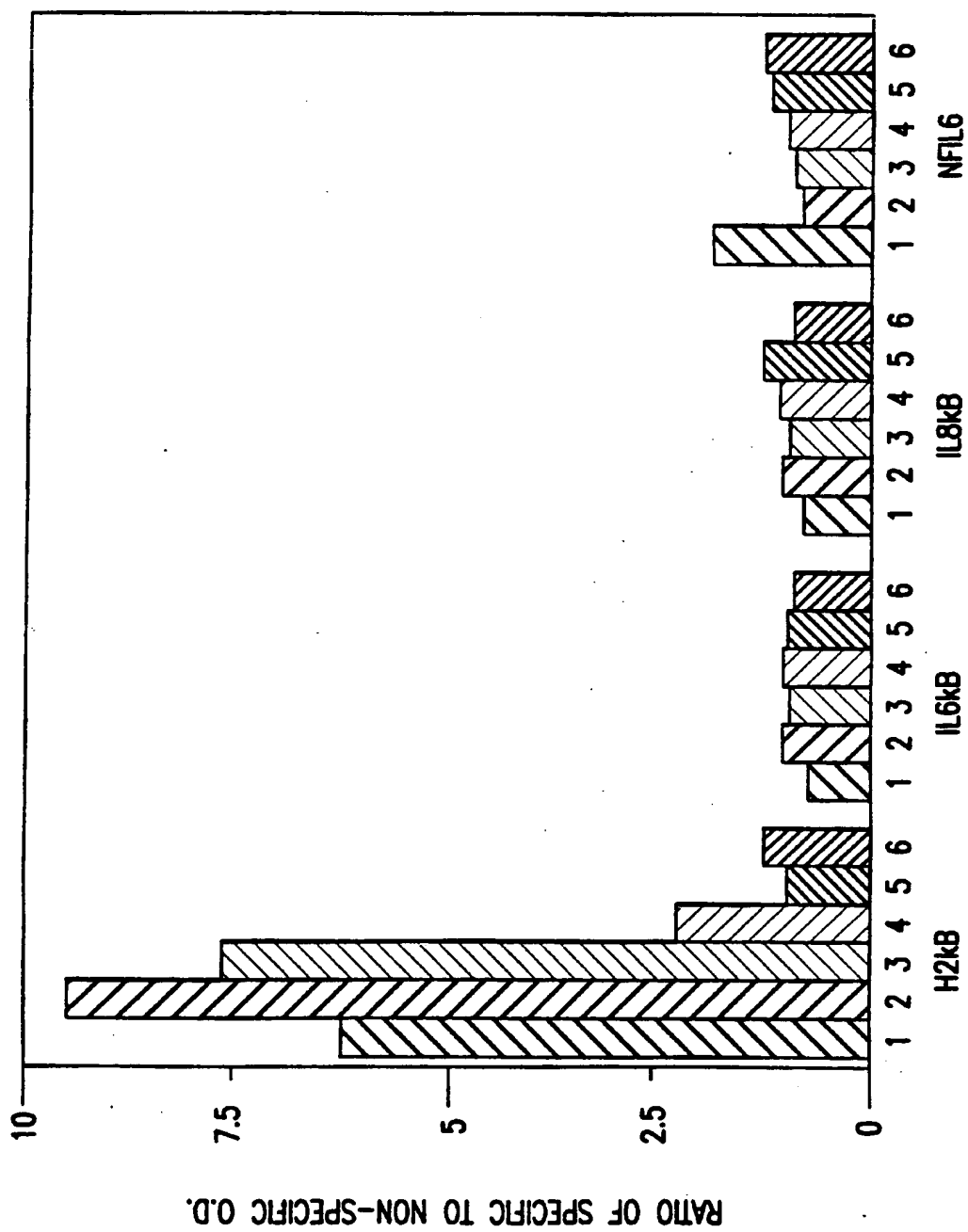


FIG.4

7/24

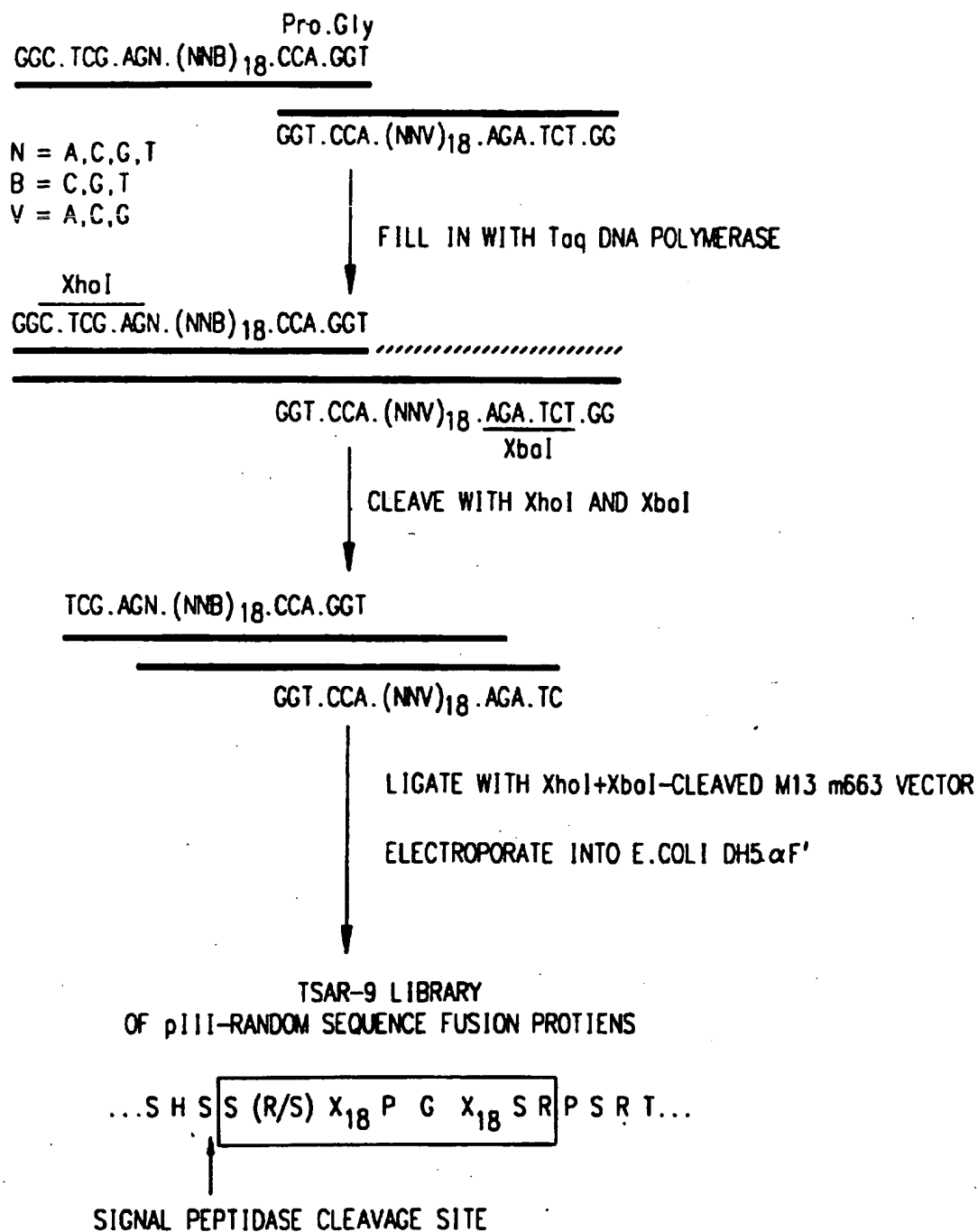


FIG.5

SUBSTITUTE SHEET (RULE 26)

8/24

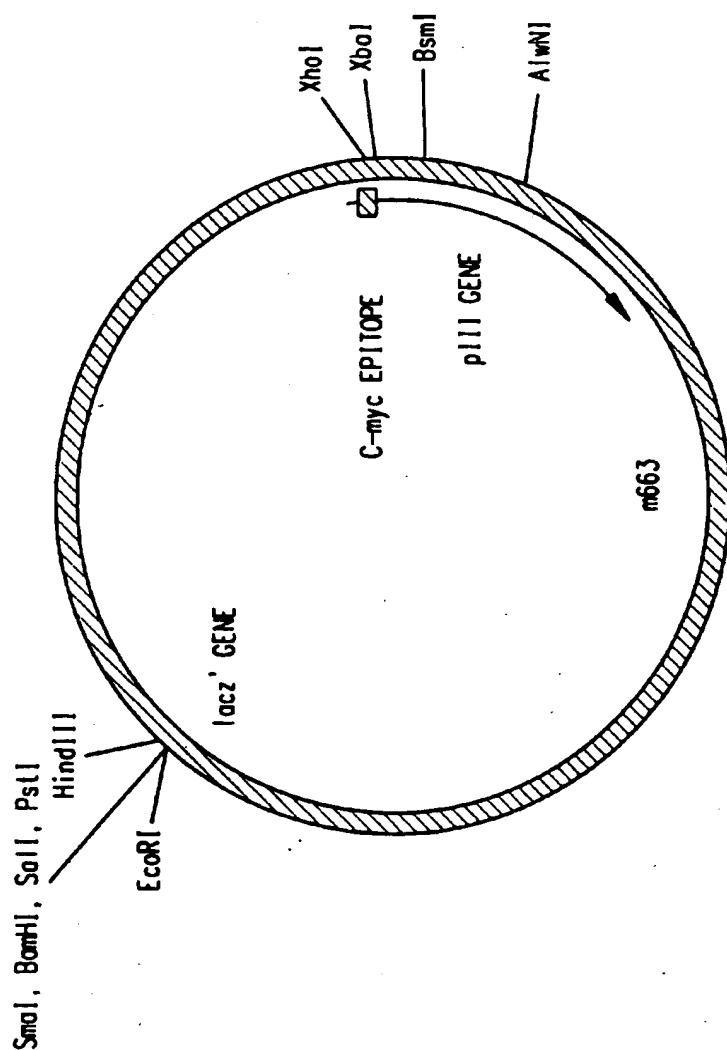


FIG. 6



9/24

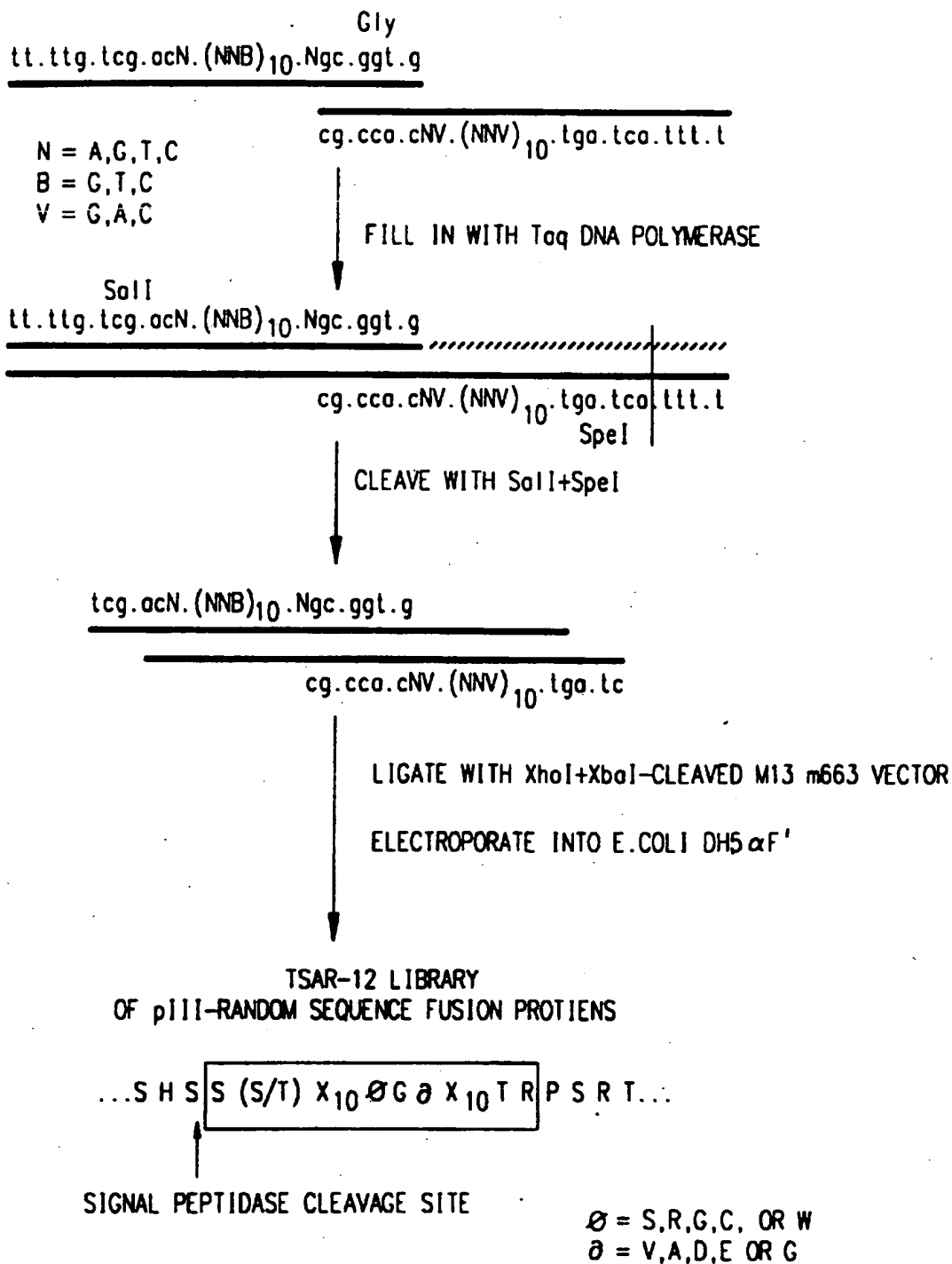


FIG.7

SUBSTITUTE SHEET (RULE 26)

10/24

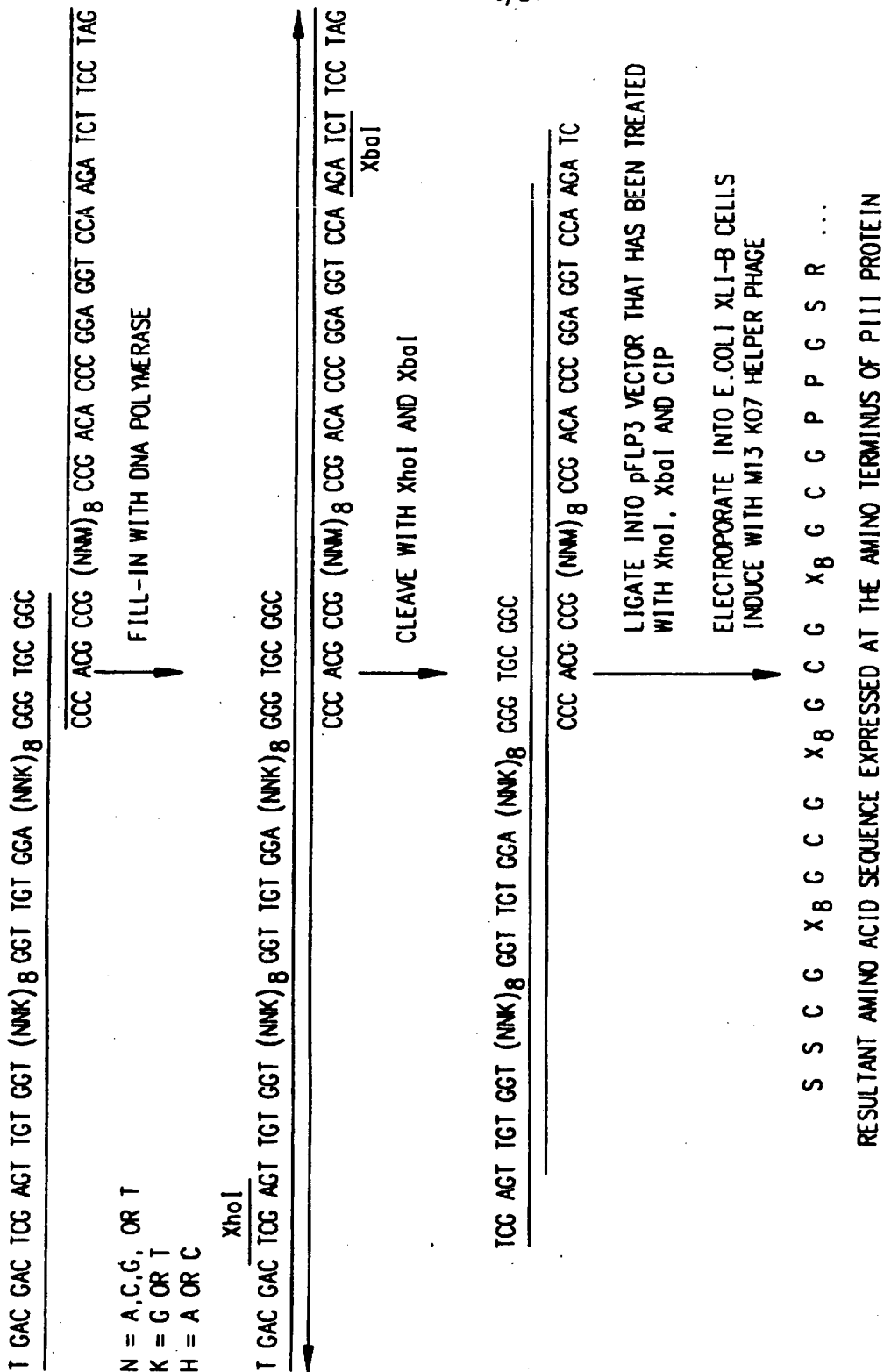


FIG.8

11/24

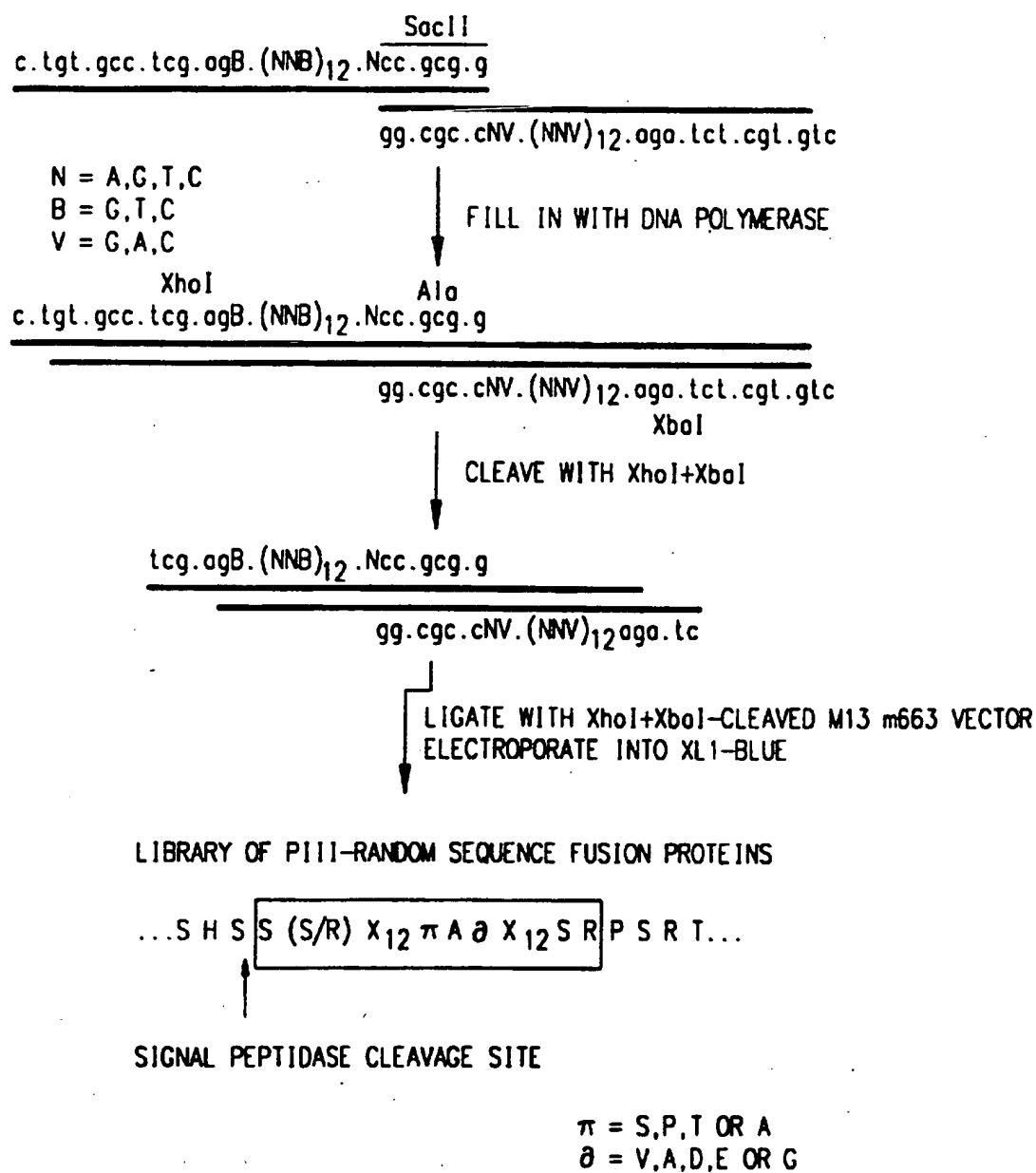
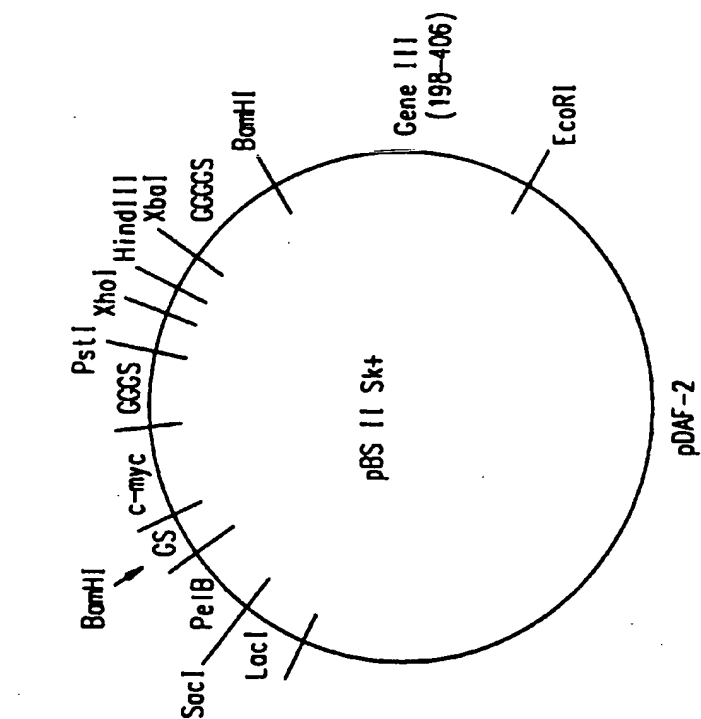


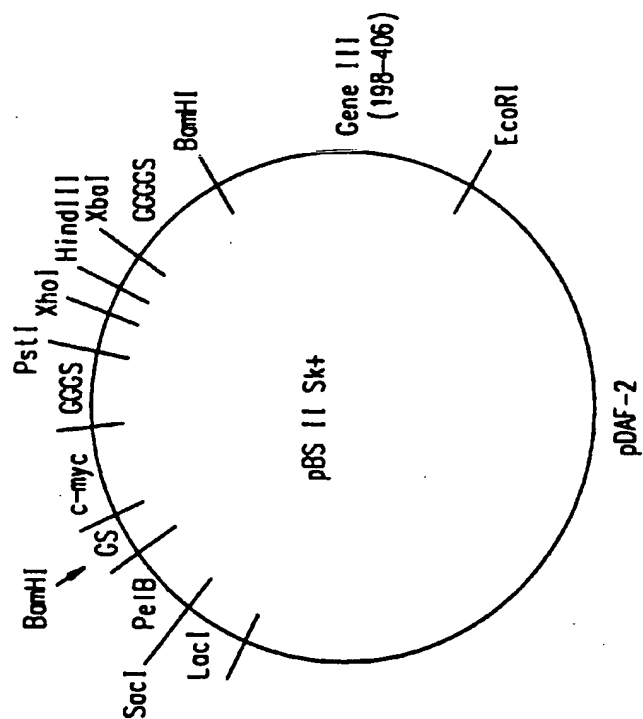
FIG.9

SUBSTITUTE SHEET (RULE 26)

12/24



**FIG. 10A**



**FIG. 10B**

13/24

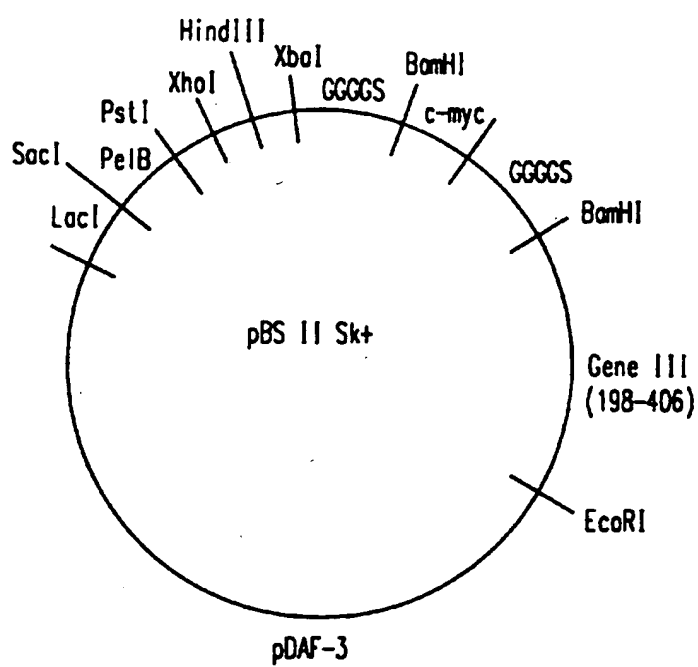


FIG.10C

14/24

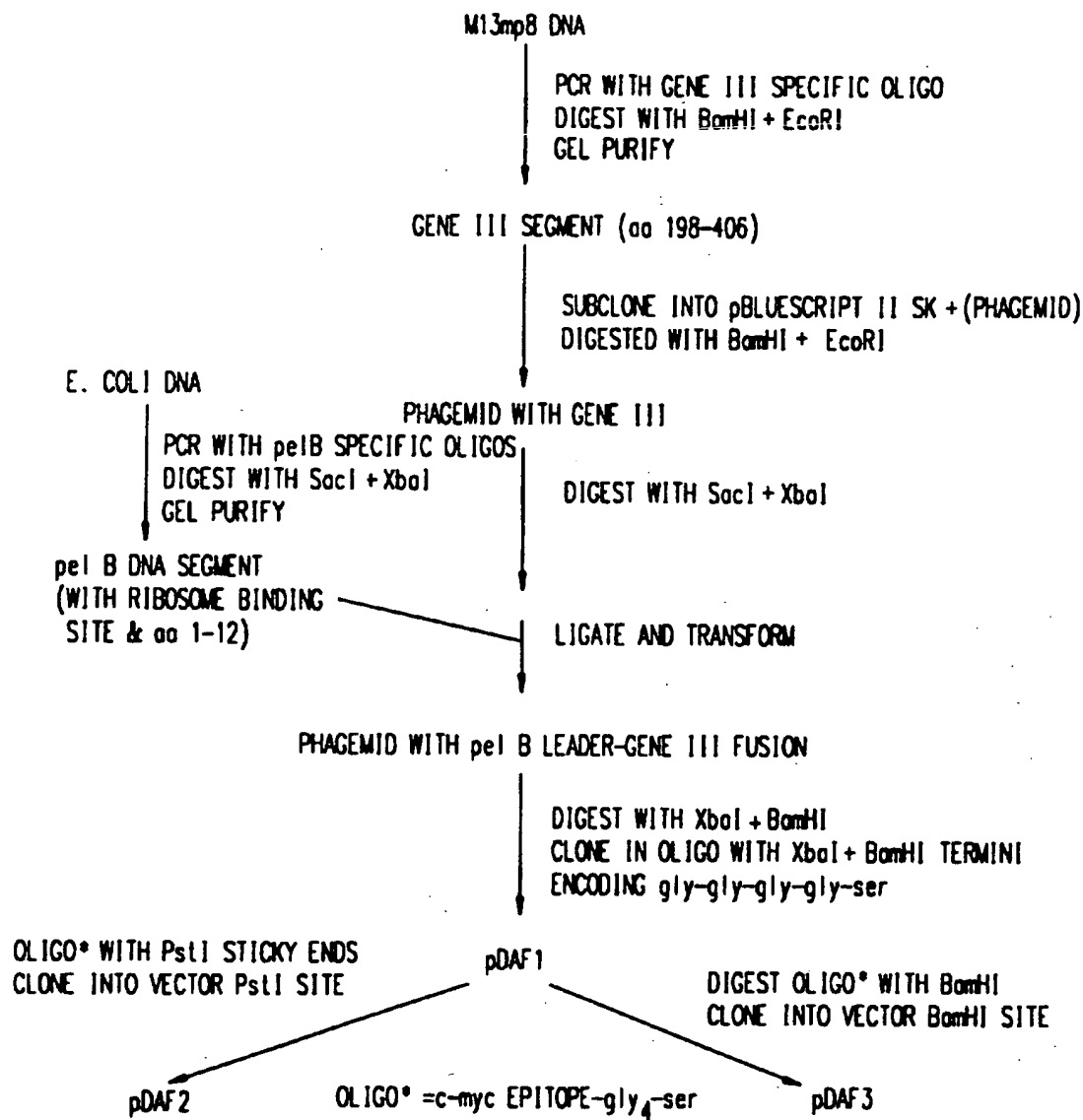
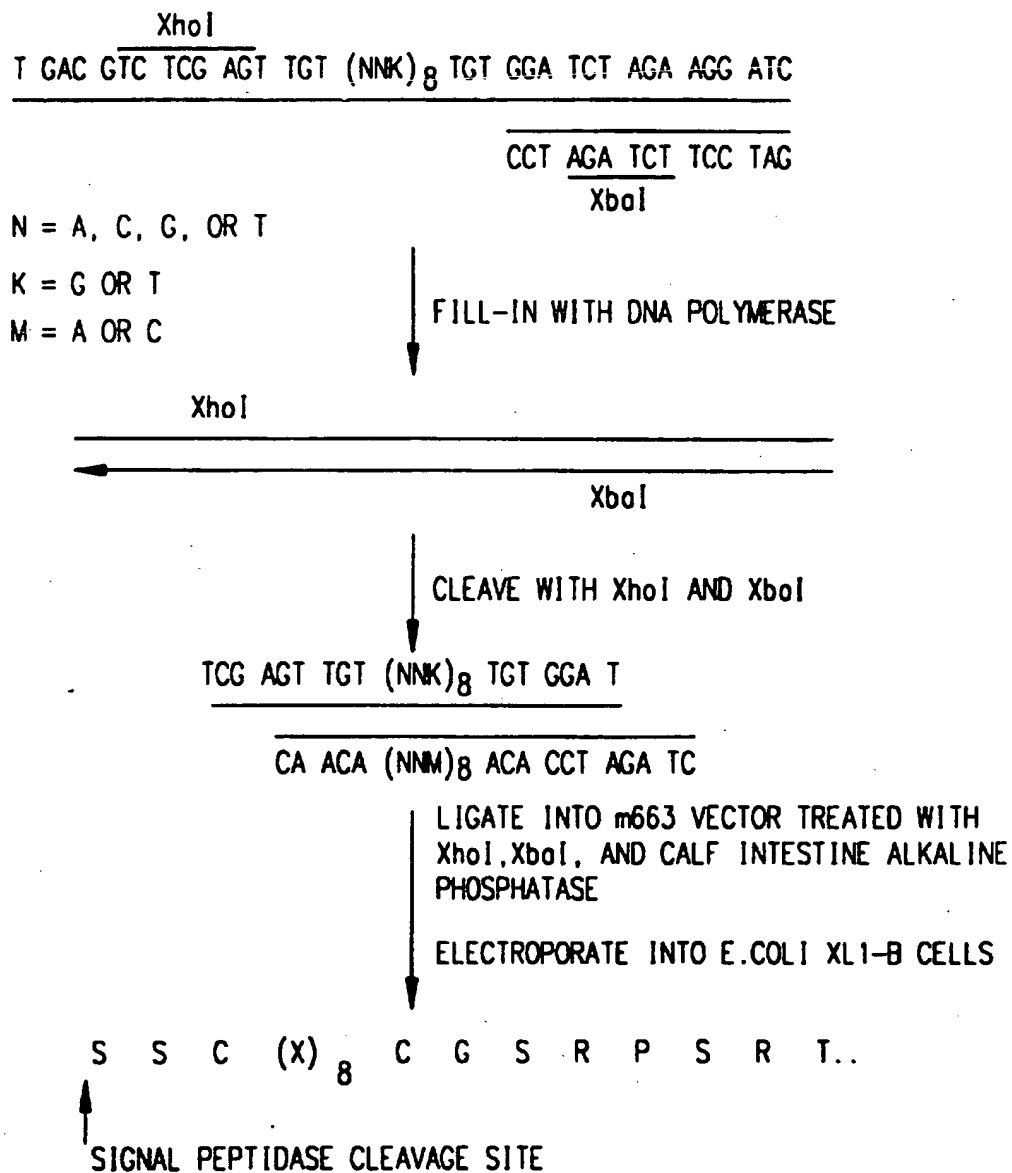


FIG.10D

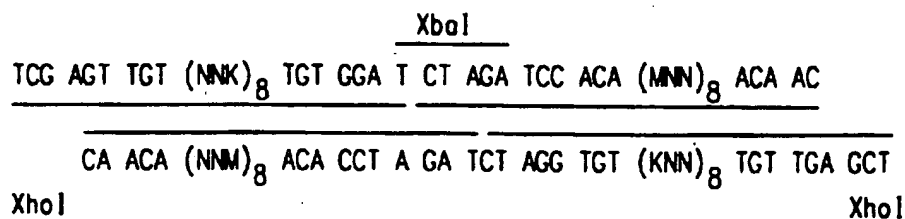
15/24



RESULTANT AMINO ACID SEQUENCE EXPRESSED AT THE AMINO TERMINUS OF p111

FIG.11

16/24



LIGATION OF TWO DOUBLE-STRANDED  
OLIGONUCLEOTIDES INTO A HEAD-TO-HEAD  
ARRANGEMENT AT THE XbaI SITE

INSERTION INTO m663 VECTOR CLEAVED  
BY XhoI AND XbaI. ILLEGITIMATE LIGATION  
AT THE XbaI SITE OF THE VECTOR.

ELECTROPORATE INTO E. COLI XL1-B CELLS

S S C (X)<sub>8</sub> C G S R S T (X)<sub>8</sub> T T R . . .

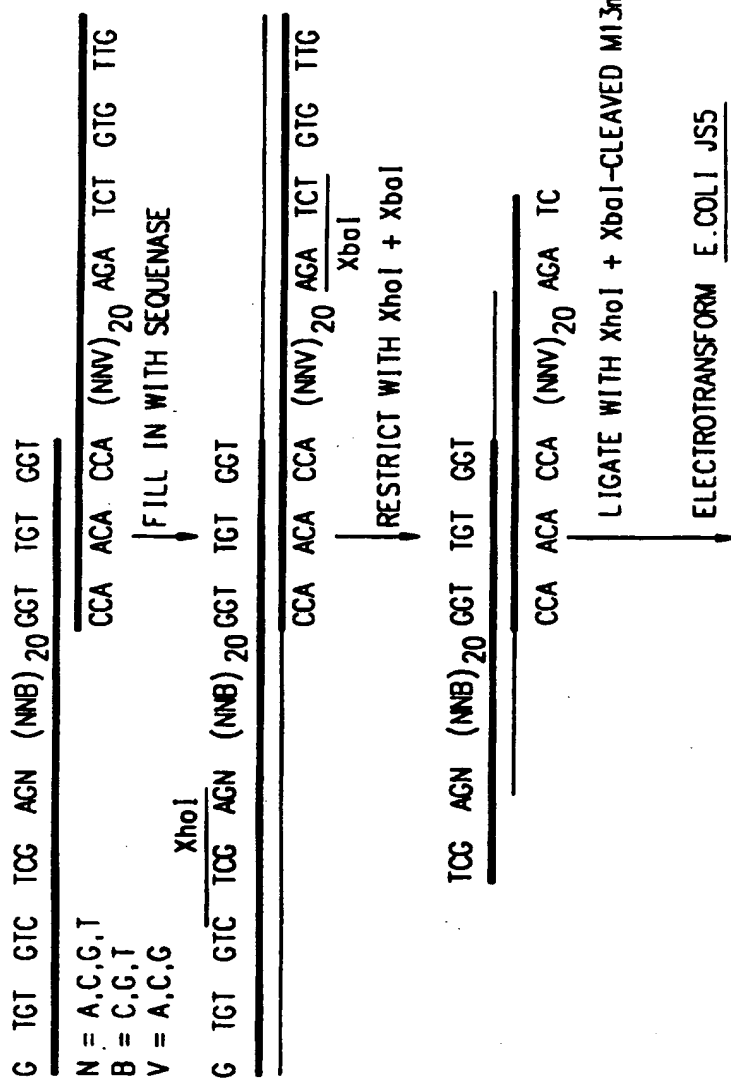
↑  
SIGNAL PEPTIDASE CLEAVAGE SITE

RESULTANT AMINO ACID SEQUENCE EXPRESSED AT THE AMINO TERMINUS OF PIII

FIG.12



17/24



DC43 GENETIC DIVERSITY LIBRARY DISPLAYED AS RANDOM N-TERMINAL pIII FUSIONS



FIG.13

18/24

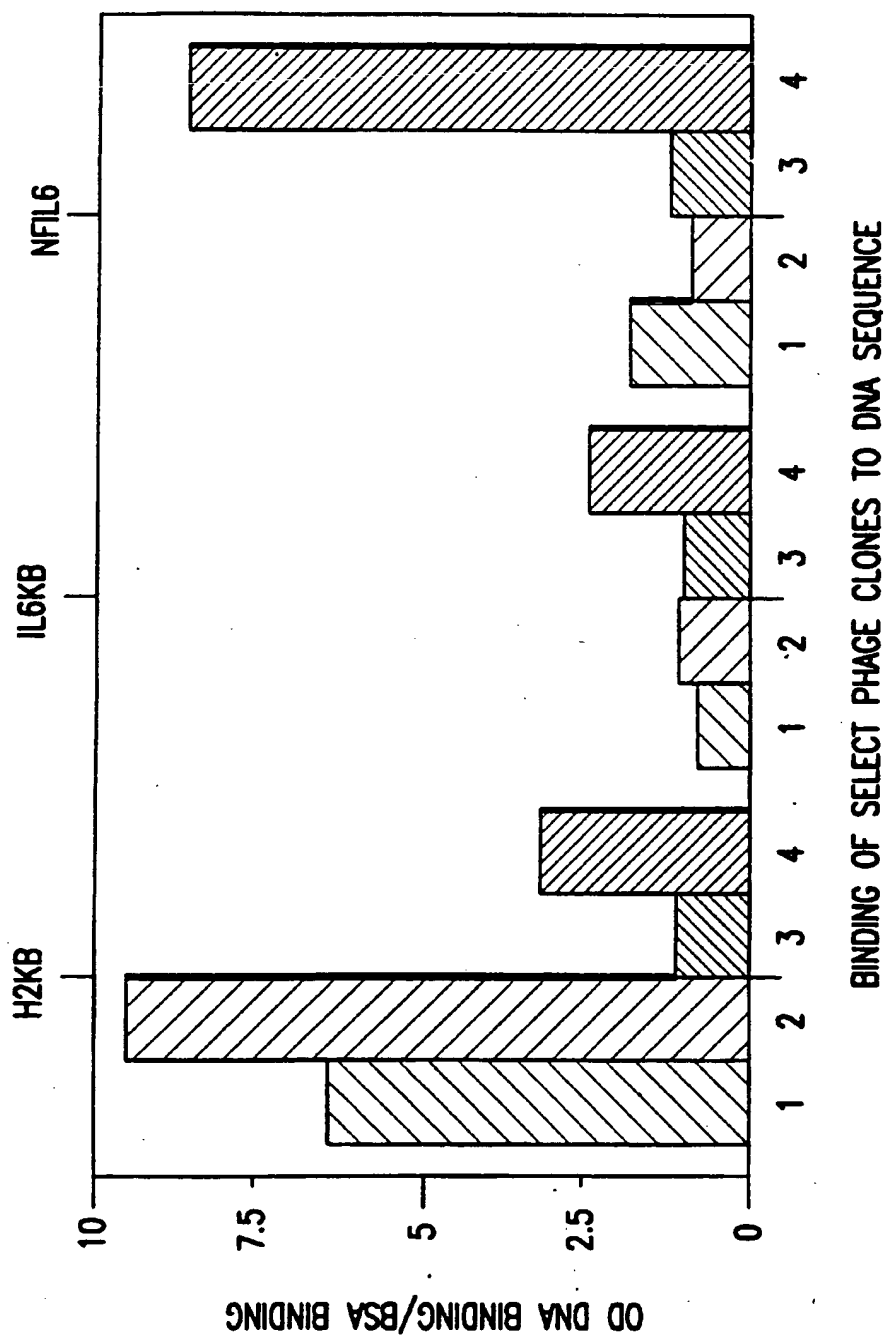


FIG.14

19/24

% 100100 94 77 77 51 25 25 77 77 77 51 51 77 51 77 25  
RESIDUE S S R T G N E Q P P G S F G R A A  
SCHEME TCCTCGAGaaccbgbgaobgabcabccbccbgbblcbltbggbcgbcnCT

%p 100 17 51 51 77 77 51 25 25 25 25 79 51  
RESIDUE G C F H P G C K Y M K L N  
SCHEME GGTtNbttbcabccbgqblqbaablbablbababctbaob

A=100%A  
G=100%G  
C=100%C  
T=100%T  
n=25%A+25%G+25%C+25%T  
a=88%A+4%G+4%C+4%T  
g=88%G+4%A+4%C+4%T  
c=88%C+4%A+4%G+4%T  
t=88%T+4%A+4%G+4%C  
b=33%T+33%C+33%G  
v=33%A+33%C+33%G

o8909:GTGTGTCTCGAGaocbggboabgabcbcbccbggblcbltbggbcgbcbnCTGGTTn 3'  
o8545:GTGTGTGTCTAGAvltlvagvllvaltvlavltlvavccvqgvltqvaavnAACCAGn 3'

08909: \_\_\_\_\_ o8545  
FILLIN WITH dNTPs AND SEQUENASE VERSION 2.0

CUT WITH XbaI AND XhoI

CLONE INTO m663 OR mBAX

FIG. 15

20/24

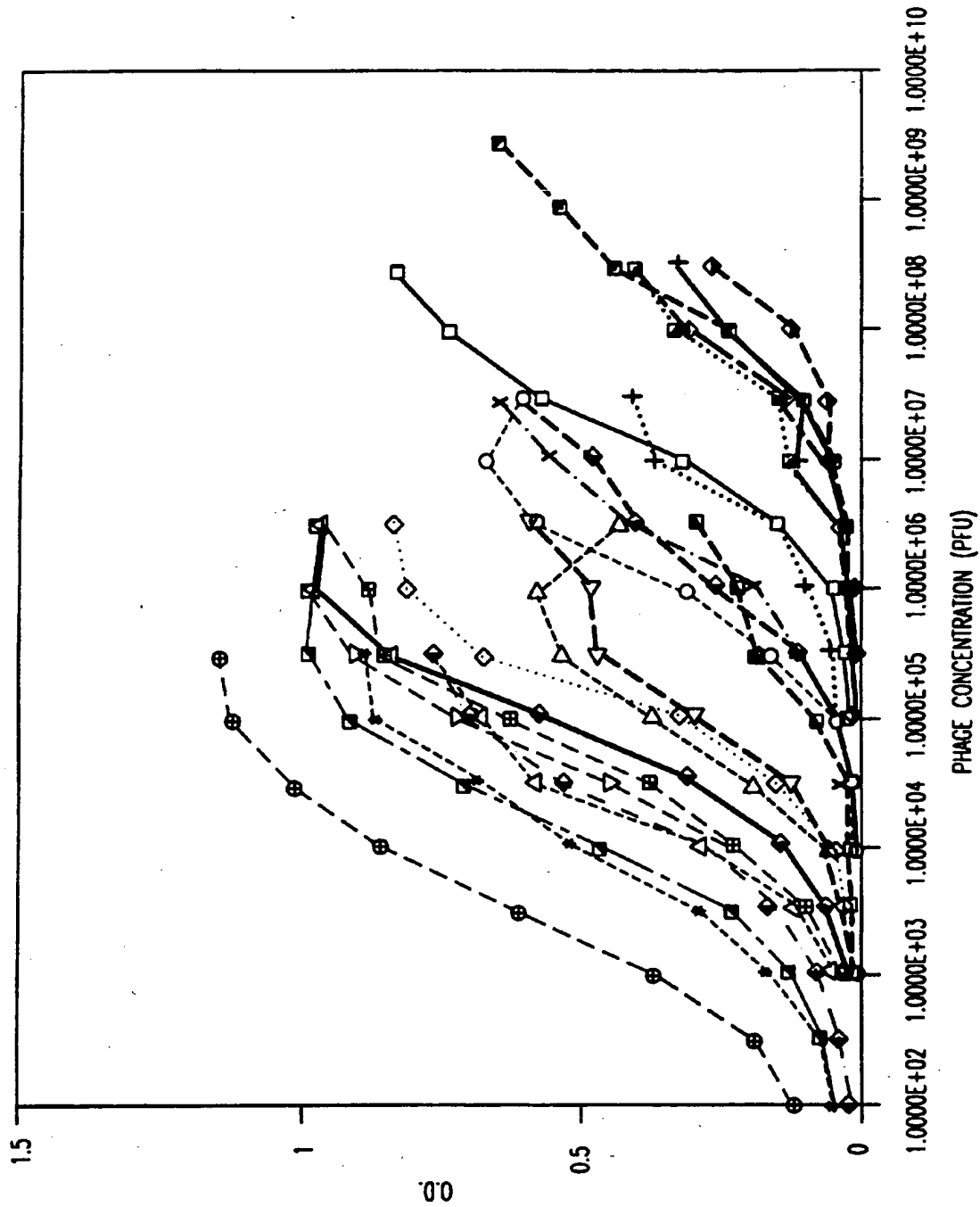


FIG.16A

21/24

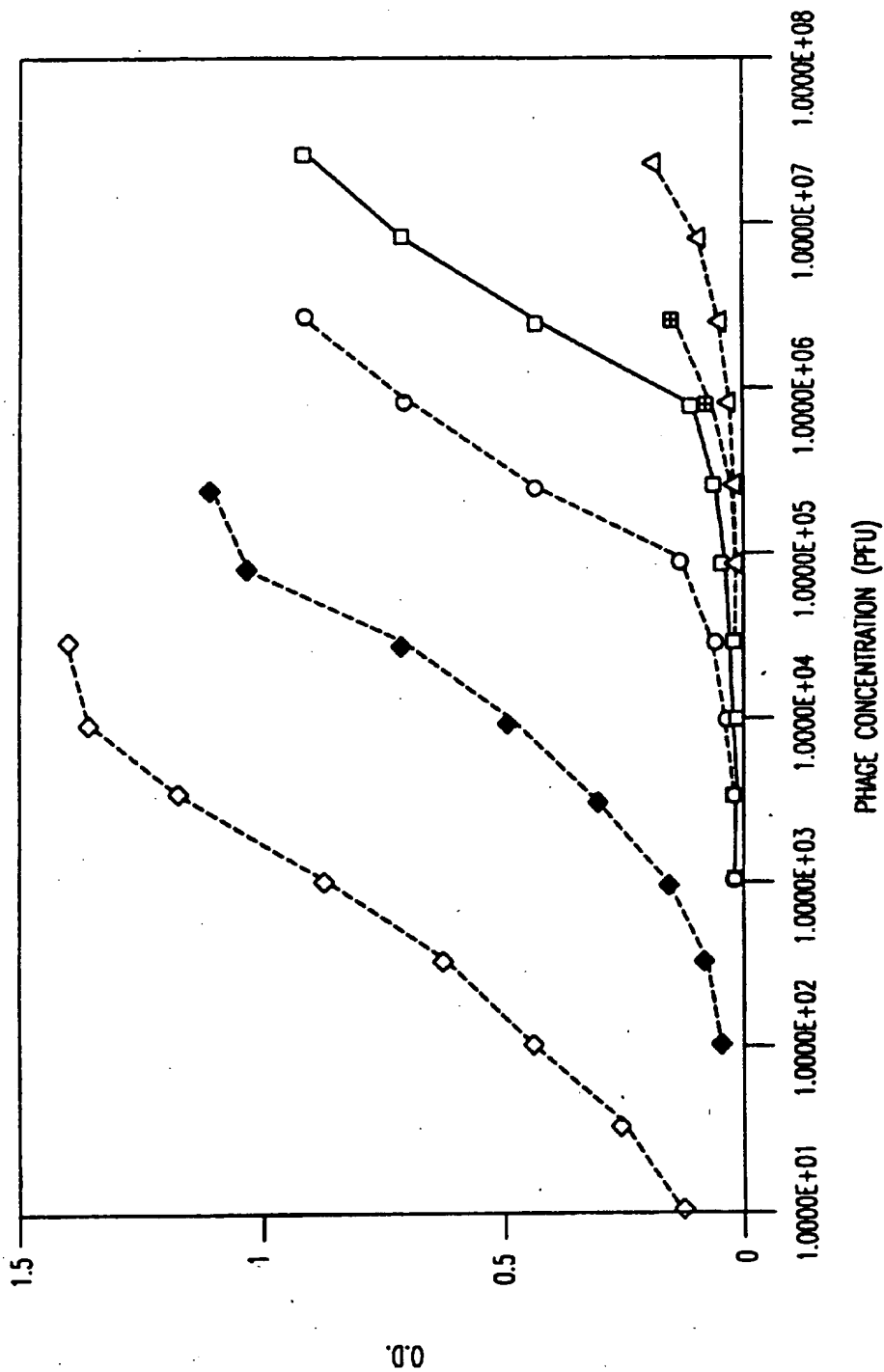


FIG.16B



23/24

X h o 	x12	x12	X b o
TGTGTCACAACCTCTCGAGNNN <del>GG</del> TTGCTTCCACCCGGTTGCAAATACATCAAACACAAGNN <del>BT</del> CTAGAACACACACA			
ACACACACAAGGAGAGCTC <del>NN</del> VCCAACGAAGGTGGCCCCAACGTTTATGTAGTTTGTGTT <del>CNN</del> VAGATCTGTGTGT			
ENCODES			
PEPTIDE: S S S/R ? G C F H P G C K Y I K H K ? S R			
x12			

## OLIGONUCLEOTIDES SYNTHESIZED:

o10180: TGTGTGTTCTAGAVNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTTGTGTTTGATGTATTTGCAACCC

o10181: TGTGTGTTCTCTCGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN~~GG~~TTGCTTCCACCCGGTTGCAAA

WHERE G=C, T=T, C=C, A=A, N=G+C+T+A, B=G+T+C, V=G+A+C

o10181: \_\_\_\_\_

o10180

FILL-IN WITH dNTPs AND SEQUENASE VERSION 2.0

CUT WITH XbaI AND XhoI

CLONE INTO m663 OR mBAX—&gt; /2b ME LIBRARY

FIG.17B

24/24

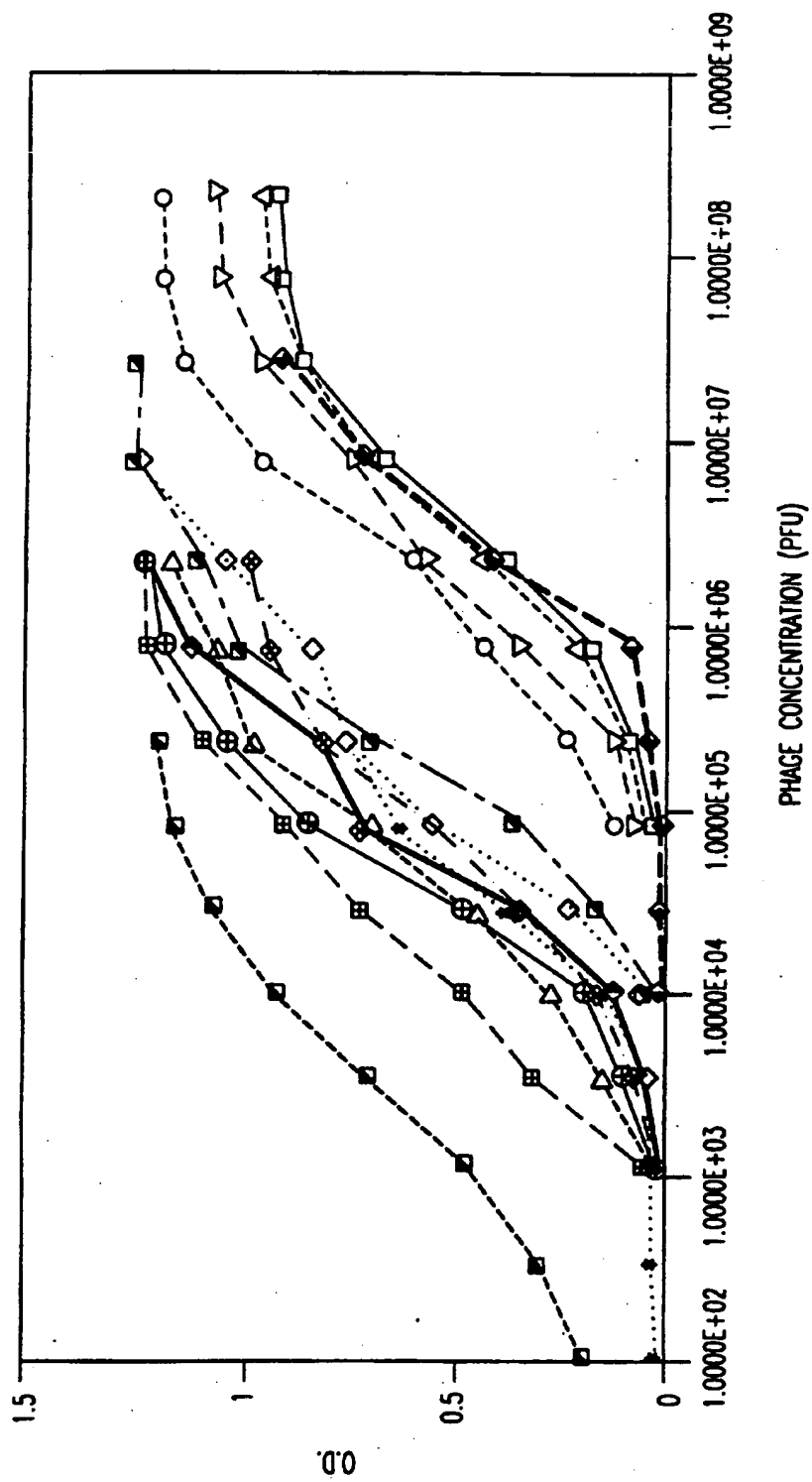


FIG.18

SUBSTITUTE SHEET (RULE 26)



## INTERNATIONAL SEARCH REPORT

 International application No.  
PCT/US95/10523
**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : C12Q 1/68; C12N 15/00; A01N 37/18; A61K 38/00, 48/00; C07H 21/02

US CL : 514/2, 44; 536/23.1; 530/300; 435/6, 172.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/2, 44; 536/23.1; 530/300; 435/6, 172.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, BIOSIS, EMBASE, MEDLINE, DERWENT, CAO

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Proceedings of the National Academy of Sciences, USA, Volume 90, issued November 1993, Matsusaka et al, "Transcription factors NF-IL6 and NF- $\kappa$ B synergistically activate transcription of the inflammatory cytokines, interleukin 6 and interleukin 8", pages 10193-10197, see the entire document.	17, 18, 32, 33, 43, 44, 60, 61, 73
Y	Molecular and Cellular Biology, Volume 13, Number 7, issued July 1993, Ko et al, "DNA-Binding specificities of the GATA transcription factor family", pages 4011-4022, see the entire document.	19, 34, 45, 62

☒ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier documents published on or after the international filing date	Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	Z	document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means		
*P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

03 NOVEMBER 1995

Date of mailing of the international search report

28 NOV 1995

 Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRUCE CAMPELL

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/10523

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Oncogene, Volume 9, issued 1994, Chang et al, "Mechanism of expression and role in transcriptional control of proto-oncogene NFKB-2/LYT-10", pages 923-933, see the entire document.	17, 18, 32, 33, 43, 44, 60, 61, 73
Y	Nature, Volume 358, issued 23 July 1992, Kim et al, "Retinoblastoma gene product activates expression of the human TGF- $\beta$ 2 gene through transcription factor ATF-2", pages 331-334, see the entire document.	19, 34, 45, 62
Y	EMBO Journal, Volume 8, Number 12, issued 1989, Risse et al, "Asymmetrical recognition of the palindromic AP1 binding site (TRE) by Fos protein complexes", pages 3825-3832, see the entire document.	19, 34, 45, 62
Y	Proceedings of the National Academy of Sciences, USA, Volume 90, issued November 1993, Needels et al, "Generation and screening of an oligonucleotide-encoded synthetic peptide library", pages 10700-10704, see the entire document.	1-73
Y	Nature, Volume 354, issued 07 November 1991, Houghten et al, "Generation and use of synthetic peptide libraries for basic research and drug discovery", pages 84-86, see the entire document.	1-73
Y	Gene, Volume 128, Number 1, issued 1993, DeGraaf et al, "Biochemical diversity in a phage display library of random decapeptides", pages 13-17, see the entire document.	1-73
Y	Science, Volume 249, issued 27 July 1990, Devlin et al, "Random peptide libraries: A source of specific protein binding molecules", pages 404-406, see the entire document.	1-73
Y	Biochemical Journal, Volume 301, issued 1994, Pinilla et al, "Investigation of antibody-antigen interactions using a soluble, non-support-bound synthetic decapeptide library composed of four trillion ( $4 \times 10^{12}$ ) sequences", pages 847-853, see the entire document.	1-73
Y	US, A, 5,338,665 (SCHATZ ET AL) 16 August 1994, see the entire document.	1-73
Y	Biochimica et Biophysica Acta, Volume 1072, issued 1991, Bauerle et al, "The inducible transcription activator NF- $\kappa$ B: regulation by distinct protein subunits", pages 63-80, see the entire document.	16, 20-22, 31, 35, 36, 42, 50, 51